Framtidens domstol

En utforskande studie kring implementering av AI inom det svenska domstolsväsendet

Kandidatarbete inom Industriell ekonomi

ANDREAS ANDERSSON
JONATAN LANDIN
TOBIAS KARLSSON
MARKUS TURESSON

INSTITUTIONEN FÖR TEKNIKENS EKONOMI OCH ORGANISATION
AVDELNINGEN FÖR ENTREPRENÖRSKAP OCH STRATEGI

CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2020
www.chalmers.se
Kandidatarbete TEKX04-20-04

Kandidatarbete TEKX04-20-04

Framtidens domstol

En utforskande studie kring implementering av AI inom det svenska domstolsväsendet

Court of the future

An explorative study on the implementation of AI within the Swedish judiciary

ANDREAS ANDERSSON
JONATAN LANDIN
TOBIAS KARLSSON
MARKUS TURESSON

TEKNIKENS EKONOMI OCH ORGANISATION
Avdelning för Entreprenörskap och Strategi
CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2020

Framtidens domstol
En utforskande studie kring implementering av AI inom det svenska domstolsväsendet

ANDREAS ANDERSSON TOBIAS KARLSSON
JONATAN LANDIN MARKUS TURESSON

Kandidatarbete TEKX04-20-04

Teknikens ekonomi och organisation
Chalmers tekniska högskola
412 96 Göteborg
Sverige
Telefon + 46 (0)31-772 1000

Göteborg, Sverige 2020
Gothenburg, Sweden 2020

Court of the future
An explorative study on the implementation of AI within the Swedish judiciary

ANDREAS ANDERSSON TOBIAS KARLSSON
JONATAN LANDIN MARKUS TURESSON

Department of Technology Management and Economics
Chalmers University of Technology

SUMMARY

This paper aims to further the knowledge regarding what possibilities exists for the implementation of AI in the Swedish judiciary, and what ramifications a potential use of AI would have for Swedish courts and society. Currently, AI is making a rapid advance in our society, AI programs are being integrated into many aspects of life. How will one of society's most slow-moving institutions, the courts, make use of the technical advances made in the fast-moving field of AI?

Today Swedish courts are not wholly uniform in their judging. Furthermore, the administrative courts have a substantial backlog of court cases. In this study, the possibility of using AI to help solve these issues is researched, and the main factors impacting the successful use of AI within the courts are identified.

The study contains a set of 16 interviews with experts on either AI or law. From the interviews different perspectives were determined and categorized into themes. The knowledge gathered from these themes were combined with a theoretical framework to identify four key factors for successfully utilizing AI in Swedish courts. These consist of building the AI for a specific purpose, clarifying the burden of responsibility, developing the AI with transparency and explainability in mind and being aware of issues stemming from bias.

Keywords: Artificial Intelligence, Judiciary, Automatization, Bias, Transparency

Note: The report is written in Swedish.

SAMMANFATTNING

Syftet med denna rapport är att undersöka möjligheterna att implementera AI i det svenska domstolsväsendet och vilka följder som en potentiell användning av AI skulle kunna ha för Sveriges domstolar och det svenska samhället. I dagsläget används AI i allt större utsträckning inom samhällets olika funktioner, men hur kommer en av samhällets mest trögrörliga institutioner, domstolsväsendet, kunna utnyttja de tekniska framsteg som gjorts inom det snabbt avancerande fältet AI?

Svenska domstolar är idag inte helt enhetliga i sitt dömande. Dessutom har förvaltningsrätterna väldigt långa handläggningstider i vissa typer av mål. I denna studie utforskas AI:s möjlighet att lösa dessa problem. Vidare identifieras

huvudfaktorerna som påverkar hur framgångsrik användningen av AI inom domstolsväsendet blir.

Studien baseras på underlag från 16 intervjuer med områdesexpert på antingen AI eller juridik. Från intervjuerna fångades olika perspektiv upp, som därefter delades in i teman. Kunskapen och informationen som samlades in från dessa teman kombinerades med ett teoretiskt ramverk för att identifiera fyra nyckelfaktorer för framgångsrikt nyttjande av AI inom svenska domstolar. Dessa är att bygga AI:n utifrån ett förutbestämt syfte, att tydliggöra ansvarsfördelningen, att utveckla AI:n med fokus på transparens och förklarbarhet och att vara medveten om och försöka minimera oönskad bias.

Nyckelord: Artificiell Intelligens, Domstolsväsendet, Automatisering, Bias, Transparens

Notera: Rapporten är skriven på svenska.

Förord

Detta kandidatarbete genomfördes under våren 2020 vid avdelningen för Entrepreneurship and Strategy på institutionen Teknikens ekonomi och organisation på Chalmers tekniska högskola och omfattar 15 högskolepoäng. Detta är för oss den avslutande delen av kandidatexamen vid industriell ekonomi.

Vi vill börja med att tacka alla de personer som tagit sig tid till att ställa upp på intervjuer, och därmed gjort denna rapport möjlig. Tack för att ni bidragit med värdefulla kunskaper och insikter inom områden som varit nya för oss. Vi vill även rikta ett tack till kursens examinator Erik Bohlin för hans engagemang i kursen.

Slutligen vill rikta ett stort tack till vår handledare Charlotta Kronblad, doktorand på avdelningen för Entrepreneurship and Strategy, vars expertis, entusiasm och breda kontaktnät har varit ovärderligt för utförandet av detta arbete.

Chalmers Tekniska Högskola
Göteborg, Sverige
14:e Maj, 2020

Innehållsförteckning

Ordlista

Allmänna domstolar - De allmänna domstolarna handlägger brottmål och tvistemål (tvister mellan enskilda). De allmänna domstolarna är tingsrätter, hovrätter och Högsta domstolen.[1]

Artificiell intelligens (AI) - Ett program som har förmågan att utföra uppgifter normalt associerade med mänsklig intelligens.[2]

Bias - Något eller någon motsätter sig eller stödjer en särskild person eller sak på ett orättvist sätt, på grund av personliga åsikter som påverkar omdömet.[3]

Brottsrubricering - Straffrättsligt namn på viss gärning, exempelvis mord, rån etcetera[4]

Dataset - En uppsättning av flera typer av fristående information, vilka behandlas som en enhet av en dator.[5]

Falskt positiv - När ett söksystem uppger att det har hittat något, trots att det inte är vad söksystemet är avsett att hitta.[6]

Förvaltningsdomstol - Förvaltningsdomstolar handlägger mål som bland annat rör förhållandet mellan enskilda individer och det allmänna, exempelvis skattemål, utlännings- och medborgarskapsmål eller socialmål.[7]

Generell AI - En AI som, likt en människa, kan hantera vilken uppgift som helst inom rimliga gränser.[8]

Hovrätt - Hovrätten är den andra instansen inom de allmänna domstolarna. Det gäller både brottmål och tvistemål som överklagats från tingsrätten.[9]

Högsta domstolen (HD) - Högsta domstolen är den tredje och sista instansen av de allmänna domstolarna. För att ett mål ska tas upp av HD krävs i nästan alla fall ett prövningstillstånd. Detta beviljas endast om målet är av betydelse för bedömning av andra mål av liknande karaktär, s.k. prejudikat.[10]

---

[1] Allt om Juridik, *Juridisk ordlista*. Hämtad 2020-03-06 från https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

[2] IT-ord, *Artificiell intelligens, Hämtad 2020-03-06* från https://it-ord.idg.se/ord/artificiell-intelligens/

[3] Cambridge Dictionary, *Bias.* Hämtad 2020-05-07 från https://dictionary.cambridge.org/dictionary/english/bias

[4] Allt om Juridik, *Juridisk ordlista* Hämtad 2020-03-06 från https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

[5] Cambridge dictionary, Dataset. Hämtad 2020-05-09 från https://dictionary.cambridge.org/dictionary/english/dataset

[6] IT-ord, *Falsk positiv.* Hämtad 2020-03-06 från https://it-ord.idg.se/ord/falsk-positiv/

[7] Allt om Juridik. *Juridisk ordlista.* Hämtad 2020-03-06 från: https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

[8] Hackernoon.com, General vs narrow AI. Hämtad 2020-05-09 från: https://hackernoon.com/general-vs-narrow-ai-3d0d02ef3e28

[9] Allt om Juridik, *Juridisk ordlista.* Hämtad 2020-03-06 från: https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

[10] Allt om Juridik, *Juridisk ordlista.* Hämtad 2020-03-06 från: https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

Juridiskt mål - Benämning på det åtal eller den tvist som är föremål för rättegång vid en juridisk tvist.[11]

Maskininlärning - Algoritmer som klarar av att lösa specifika uppgifter utan att använda explicita instruktioner, istället används mönster och slutledningsförmåga för att producera resultat.[12]

Rättssäkerhet - Saknar entydig definition, men innebär att systemet ger den enskilda medborgaren skydd för godtyckliga ingrepp från samhället självt, att domar inte faller utan tydligt lagstöd och tydlig bevisning, samt att alla medborgare bedöms på ett likartat sätt.[13]

Snäv AI - AI som är specialiserad på att hantera en enda uppgift.[14]

Tingsrätt - Tingsrätten är den första instansen inom det allmänna domstolsväsendet. En tingsrätt prövar brottmål och tvistemål.[15]

---

11 Allt om Juridik, *Juridisk ordlista.* Hämtad 2020-03-06 från https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

12 Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A, *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9, Hämtad 2020-03-06

13 Allt om Juridik, *Juridisk ordlista.* Hämtad 2020-03-06 från https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

14 DeepAi.org, What is Narrow AI?. Hämtad 2020-05-09 från https://deepai.org/machine-learning-glossary-and-terms/narrow-ai

15 Allt om Juridik, *Juridisk ordlista.* Hämtad 2020-03-06 från https://www.alltomjuridik.se/lar-dig-juridiska/juridisk-ordlista/

1. Inledning

I kapitlet *Inledning* presenteras bakgrunden till studien och vilka problem den avser att utforska, dessa problem analyseras och utifrån detta formuleras studiens syfte och frågeställning. Slutligen redogörs för studiens avgränsningar.

1.1 Bakgrund

Att alla människor är lika inför lagen är en stöttepelare i en väl fungerande rättsstat. Denna princip går under namnet likhetsprincipen. Principen är fastlagd i svensk lag, "Domstolar samt förvaltningsmyndigheter och andra som fullgör offentliga förvaltningsuppgifter ska i sin verksamhet beakta allas likhet inför lagen samt iaktta saklighet och opartiskhet" (Lag, SFS 2010:1408). Enligt Förenta nationerna (FN, 2008) är likhet inför lagen och juridiskt skydd utan diskriminering en mänsklig rättighet. Svensk lag fastställer också att en rättegång måste hållas inom skälig tid, "En rättegång ska genomföras rättvist och inom skälig tid. Förhandling vid domstol ska vara offentlig" (Lag, SFS 2010:1408).

En studie från Brottsförebyggande rådet (Brå, 2017) fann att det existerade ett antal betydande ojämlikheter inom det svenska domstolsväsendet. Det fanns bland annat utbredda geografiska skillnader, till exempel var sannolikheten att dömas till fängelse i Lycksele tingsrätt mindre än hälften jämfört med i Värmlands tingsrätt. I ett annat exempel från rapporten fokuserades särskilt på grovt rattfylleri. Brå identifierade att det var 60 procent större sannolikhet att män dömdes till fängelse för förseelsen jämfört med kvinnor, samt att det geografiskt fanns stora skillnader med avseende på påföljd. I Lycksele dömdes endast 34 procent till fängelse, motsvarande siffra uppgick i Örebro tingsrätt till 62 procent. Rapporten nämner naturliga förklaringar till detta, som att individerna som dömdes i Lycksele i större utsträckning var under 21 år och att de dömda i Örebro redan existerade i brottsregistret. Dock understryks i rapporten att spridningen, dessa faktorer till trots, var anmärkningsvärd.

Enligt Regeringskansliet (2019) har antalet inkomna ärenden till Sveriges domstolar ökat markant under de senaste åren. I syfte att bibehålla domstolarnas höga kvalitet och effektivitet föreslår regeringen att det ekonomiska stödet till Sveriges domstolar bör ökas under de kommande åren. Under 2020 respektive 2021 vill regeringen öka domstolarnas

anslag med 280 respektive 275 miljoner kronor, detta anslag var under 2019 knappt 5,7 miljarder kronor (Regeringen, 2018). Utöver de ökade kostnaderna har domstolarna även långa handläggningstider. Förvaltningsrätten i Göteborg (2019) har till exempel i mål som berör beskattnings- och socialförsäkringsärenden handläggningstider på 13–17 månader. Ännu mer utdragna processer har mål om uppehållstillstånd på grund av skydds- eller asylskäl, med handläggningstider på 20–24 månader. Även förvaltningsrätten i Linköping (2019) och i Malmö (2019) har en liknande situation rörande handläggningstider.

1.1.1 Artificiell intelligens och juridik

Begreppet artificiell intelligens (AI) myntades av professor John McCarthy år 1955 och det syftar till vetenskapen och tekniken att konstruera intelligenta maskiner och mer specifikt intelligenta datorprogram (Stanford University, u.å.). Häggström (2019) skriver att AI-forskningen befinner sig i stark expansion och att den visar stora framgångar inom ett flertal olika områden och att det eventuellt kommer ske ett större genomslag för AI-tekniker.

Att använda sig av artificiell intelligens inom juridiken är inte en ny idé. Under 1980-talet diskuterades konceptet flitigt och år 1987 hölls den första internationella konferensen som innefattade juridik och AI (Rissland, Ashley & Loui, 2003). Inom andra områden används redan AI för att fatta beslut. Ett exempel på detta är den så kallade Trelleborgsmodellen, där socialtjänsten använder algoritmer som tar fram beslutsunderlag samt beslutar huruvida en ansökan om försörjningsstöd ska beviljas eller ej (Trelleborgs kommun, 2019).

Richard Susskind och Phillip Capper utvecklade 1988 det första kommersiella AI-systemet inom juridik (Susskind, 2019). Systemet var designat att svara på frågan, “När kan en specifik handling inte längre kan göras gällande på grund av att den är preskriberad?”. Programmet var designat som ett stort beslutsträd med över två miljoner vägar, en för varje typ av fall som Susskind och Capper kunde komma på. Detta tidiga AI-program reducerade undersökningstiden i vissa fall från timmar till minuter. Susskind nämner vidare att det idag existerar en rad AI-program baserade på maskininlärning som med hjälp av stora mängder data kan identifiera mönster, likheter och samband som mänskliga advokater inte klarar av med konventionella metoder. Dessa program är enligt Susskind redan idag mer effektiva än juniora advokater gällande granskning av stora mängder text samt på att välja ut vilka dokument som är mest relevanta.

1.2 Problem och problemanalys

Sveriges domstolar arbetar idag efter likhetsprincipen - att lika fall skall behandlas lika. Brå:s rapport “Enhetligt dömande i tingsrätter” (2017) visar på en inkonsekvent bedömning av lika brott mellan Sveriges olika tingsrätter. Rapporten visar även på att sannolikheten att gärningspersonen döms till fängelse varierar beroende på kön. Dessa två faktorer pekar på att likhetsprincipen idag inte tillämpas som det är skrivet i svensk lag och att det bör ske någon form av standardisering av Sveriges domstolar. Även de svenska förvaltningsdomstolarna står inför stora utmaningar då handläggningstiderna för vissa typer av mål är mycket långa. Detta kan vara anledning att ifrågasätta om mål tas upp inom skälig tid.

Ur ett etiskt perspektiv, specifikt för tillämpningen av rättvisa, så har AI möjligheten att stärka nuvarande system. Future of Life Institute (2017) har tagit fram en lista med 23 principer för utvecklingen av AI. En av dessa principer är juridisk transparens, vilket innebär att ett autonomt system inblandat i juridiskt beslutsfattande ska kunna förklara sina resultat till en kompetent mänsklig auktoritet. Att AI ska designas och användas så att den korrelerar med mänskliga värderingar angående värdighet, rättigheter, friheter och kulturell mångfald är också en av dessa principer. Future of Life Institute nämner även andra principer för AI-utveckling, som syftar till att så många som möjligt ska kunna dra nytta av AI och välståndet som skapas. Dessa principer beskriver inte bara utmaningar gällande utvecklingen av AI utan även möjligheter till förbättringar för hela mänskligheten. Många av principerna kan om de följs bidra till att FN:s globala mål för hållbar utveckling uppnås (UNDP, 2015).

De svenska domstolarna tillämpar alltså inte likhetsprincipen fullt ut och är, speciellt gällande förvaltningsdomstolarna, inte tillräckligt effektiva. AI har eventuellt potentialen att lösa båda dessa utmaningar eftersom tekniken har stora möjligheter att både standardisera och accelerera svenska domstolars arbete.

1.3 Syfte

Studien ämnar utreda hur AI kan användas i svenska domstolar, vilka möjligheter, risker och begränsningar som finns. Resultatet kommer i sin tur att användas för att ta fram rekommendationer för implementering av AI inom det svenska domstolsväsendet.

1.3.1 Frågeställning

Studien har utgångspunkt i följande frågeställningar:

- Hur kan AI användas i svenska domstolar?
- Vilka möjligheter, risker och begränsningar finns med AI-implementering inom domstolsväsendet?

1.4 Avgränsningar

För att kunna utföra relevanta intervjuer och sammanställa adekvat litteratur avgränsades studien. Studien fokuserade på svenska domstolar och hur AI kan implementeras i dessa för att öka deras effektivitet, kvalitet och enhetlighet. Ytterligare en avgränsning var att studien inte gick in på djupet av tekniken bakom AI, utan tog ett mer översiktligt grepp.

2. Metod och genomförande

Inom ramen för kandidatarbetet har en studie genomförts som bedömdes vara explorativ, kvalitativ och av abduktiv karaktär. I syfte att få insikt i hur AI kan användas i domstolsväsendet och vad en sådan användning kan innebära, genomfördes kontinuerliga intervjuer med personer som besitter särskild områdesexpertis. Utöver detta tillkom information som hämtats i syfte att bygga upp det teoretiska ramverk som skapade underlag för diskussion.

En explorativ studie syftar till att svara på frågor som nyttjas vid bestämning av fokuspunkter för framtida undersökningar och för att få fundamentala kunskaper kring basala frågor för det studerade problemet. Eftersom studien använder sannolika samband och rimliga antaganden för att identifiera problemet, snarare än att säkert veta dess exakta beståndsdelar, nyttjade studien per definition en abduktiv metodik (Wallén, 1996). Studiens kvalitativa karaktär lämpade sig för frågeställningen då en öppnare frågeställning med områdesexperter leder till svar som är mer målande och förklarande i sin natur, samt att svaren eventuellt är oväntade och leder till nya diskussionsområden (Family Health International, u.å.).

2.1 Processens utformning

Studien baserades på datainsamling från intervjuer med områdesexperter inom såväl AI som juridik. För att underbygga vidare analys av detta underlag har ett teoretiskt ramverk skapats utifrån relevant litteratur. Därmed baseras rapporten på primärdata, med sekundärdata som stöd. Arbetet grundar sig huvudsakligen på kvalitativa data, som sammanställts och analyserats på ett för projektet relevant sätt. För att denna analys skulle vara möjlig behövde intervjuerna byggas upp systematiskt och enhetligt i den mån det var genomförbart. Samtliga intervjuer hade således utgångspunkt i samma originalmall, med viss anpassning för respektive område. Intervjuerna genomfördes med semistrukturerad intervjuteknik, som säkrar jämförbarhet mellan intervjuerna tack vare en gemensam struktur, men bibehåller möjlighet till hög flexibilitet (Bryman & Bell, 2015). Vid varje intervjutillfälle närvarade två personer, kombinerade i olika konstellationer i syfte att maximera variationen avseende perspektiv. I inledningen av samtliga intervjuer sammanfattades dess syfte, och respondentens inställning till inspelning och anonymitet kontrollerades.

2.2 Datainsamling

Under arbetsprocessens gång genomfördes flertalet intervjuer med personer som besitter särskild kunskap inom områden som berörs av arbetet. I planeringsstadiet bedömdes dessa dels vara personer aktiva inom det juridiska väsendet, och dels personer som forskar på eller arbetar med AI i allmänhet samt kombinationen AI och etik i synnerhet. Kontakt etablerades i vissa fall efter tips från handledare, i andra fall genom kontakt via e-post som tagits efter att personen bedömts vara av intresse för studien. Ett så kallat snöbollsurval användes i enstaka fall, där primära kontakter förmedlar information och intervjuare vidare till andra personer inom deras sociala nätverk (Bryman & Bell, 2015). Urvalsmetodiken kan med fördel användas för att nå ut till grupper eller individer som inte skulle nås med andra urvalsstrategier (Family Health International, u.å.). Dessutom medför nyttjandet av sociala nätverk en möjlighet att utforska de olika respondenternas relation och interaktion med varandra, samt att identifiera dynamiken i dessa nätverk (Noy, 2008). En nackdel med ett sådant respondenturval är att det är svårt att hävda generalitet. Detta eftersom valet av personen som får snöbollen att rulla är högst subjektivt och medför att urvalet är att betrakta som partiskt (Griffiths et al, 1993).

Intervjuerna genomfördes sedan i form av fysiska möten eller på distans, där majoriteten har utförts med hjälp av videosamtal men i enstaka fall utan video. Dessa samtal spelades in förutsatt godkännande från respondent som även fick ge sitt godkännande till att bli omnämnd och citerad i rapporten. För att säkerställa att respondenterna ansåg sig korrekt citerade skickades deras citat med tillhörande brödtext ut via e-post för revidering och godkännande innan publicering. Intervjuer med 16 personer, fördelade på områdena AI respektive juridik, genomfördes i syfte att ge studien önskat omfång. I enlighet med vad Guest, Bunce och Johnson (2006) belyser i sin artikel, bedömdes detta antal intervjuer ge ett informationsunderlag som motsvarar tillräcklig grad av datamättnad, vilket innebär att ytterligare intervjuer inte skulle tillföra tillräckligt mycket information för att det skulle anses vara värdefullt att utföra dem (Faulkner & Trotter, 2017).

Tabell 1: Intervjuobjekt inom AI.

Namn	**Befattning**	**Tid**
Jacob Dexe	*Industridoktorand, RISE och Kungliga Tekniska högskolan*	*44 min*
Frank Dignum	*Professor, Institutionen för datavetenskap, Umeå universitet*	*25 min*
Karl de Fine Licht	*Lektor i teknik och etik, Teknikens ekonomi och organisation, Chalmers tekniska högskola*	*49 min*
Fredrik Heintz	*Biträdande professor, Institutionen för datavetenskap, Linköpings universitet*	*57 min*
Fredrik Johansson	*Forskarassistent på avdelningen för Data science och AI, institutionen för data- och informationsteknik, Chalmers tekniska högskola*	*63 min*
Josefin Rosén	*Chefsrådgivare inom Advanced Analytics and AI, SAS Institute*	*44 min*

Thomas Schön	*Professor i reglerteknik, institutionen för Informationsteknologi, Uppsala universitet*	*44 min*

Tabell 2: Intervjuobjekt inom juridik.

Namn	**Befattning**	**Tid**
Stanley Greenstein	*Universitetslektor, Juridiska Institutionen, Stockholms universitet*	*i.u.**
Anders Hagsgård	*Hovrättspresident, Hovrätten för Västra Sverige*	*36 min*
Magnus Ivarsson	*Verksamhetsutvecklare, Domstolsverket*	*26 min*
John Lagström	*IT-strateg, Domstolsverket*	*56 min*
Anna-Sara Lind	*Professor i offentlig rätt, Juridiska institutionen, Uppsala universitet*	*29 min*
Andrea Lindblom	*Administrativ chef, Helsingborgs tingsrätt*	*36 min*
Markus Naarttijärvi	*Universitetslektor, Juridiska institutionen, Umeå universitet*	*35 min*
Johan Sanner	*Chefsrådman, Förvaltningsrätten i Göteborg*	*64 min*
Björn Östman	*Domare, Hovrätten för Västra Sverige, IT-utvecklare, Domstolsverket*	*65 min*

*Intervjun kunde inte spelas in på grund av tekniska problem.

Syftet med dessa intervjuer var, utöver att samla in primärdata till den slutliga rapporten, också att skapa förståelse för de olika parternas perspektiv på implementeringen av AI inom domstolsväsendet. För att undvika misstaget att påbörja datainsamlingen för tidigt konstruerades merparten av det teoretiska ramverket innan intervjuerna med avsikt att tydliggöra syftet (Lantz, 2013).

2.3 Tematisering och analys

I syfte att identifiera gemensamma mönster och teman för intervjuerna genomfördes en tematisk analys, vilket lämpar sig väl för att analysera kvalitativa data (Blomkvist, 2014). Eftersom tolkningar av kvalitativa data är att betrakta som subjektiva, så tolkade och analyserade samtliga gruppmedlemmar gemensamt allt intervjuinnehåll. Baserat på frågeställningen identifierades tre huvudteman, vilka sedan delades in i underteman. Dessa underteman bestämdes utifrån frekvent förekommande och relevanta begrepp, relaterade till respektive huvudtema, som framkom under intervjuerna. Detta följdes av att intervjuerna analyserades ytterligare en gång med utgångspunkt i dessa underteman. I enstaka fall bedömdes aspekter som inte var lika frekvent förekommande som särskild intressanta, dessa skapade då underteman som kompletterade den mer traditionella tematiska analysen. På grund av det subjektiva urvalet av respondenter adderades ytterligare ett huvudtema, *Respondenternas inställning*, för att tydliggöra respondenternas ståndpunkt.

I enlighet med vad Blomkvist (2014) understryker genomfördes en reduceringsprocess på samtliga teman, där innehåll som bedömdes irrelevant för vidare analys exkluderades. Denna process skedde först då gruppen genom skapandet av det teoretiska ramverket hade tillräckligt kunskapsunderlag för att bedöma diverse uppgifters relevans. Utöver detta används citat från intervjuerna för att ytterligare stärka och nyansera den sammanställning som gjorts av respondenternas svar.

3. Teoretiskt ramverk

I kapitlet *Teoretiskt ramverk* presenteras litteratur som är relevant för rapporten. Ramverket är tänkt att ge läsaren en djupare teoretisk förståelse, samt att bidra med ytterligare diskussionsunderlag till rapportens resultat.

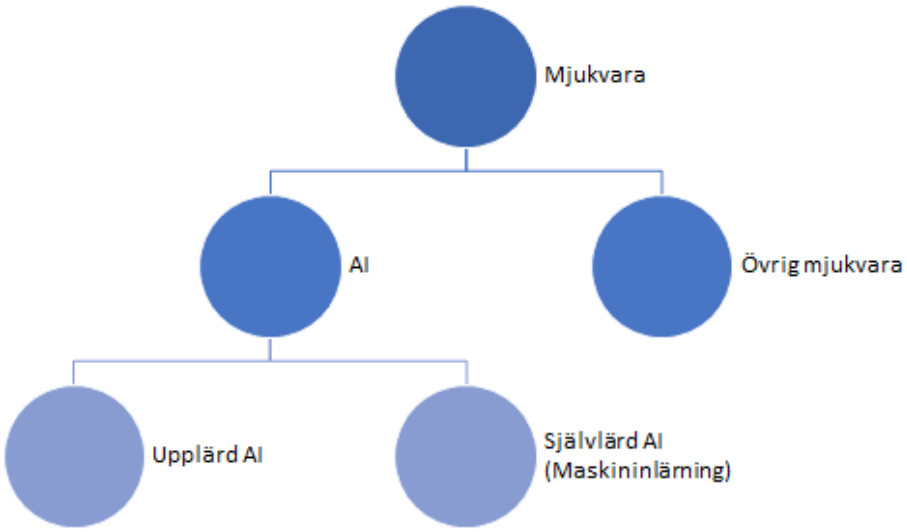
3.1 AI

Artificiell intelligens (AI) är ett begrepp som nämns i en mängd olika sammanhang, tekniken har haft en snabb tillväxt och implementeras i allt från tandborstar till självkörande bilar. Definitionen för vad AI egentligen är har varierat med tiden, men enligt

Nationalencyklopedin är syftet med AI att artificiellt kunna efterlikna den mänskliga hjärnans olika förmågor kring till exempel problemlösning, informationsinsamling och språk (Artificiell intelligens, 2020).

Att specificera vad som är och vad som inte är AI är svårt. Häggström (2019) menar att många aktörer på senare år har valt att räkna sina egna system som AI för publicitet och finansiering och att detta har bidragit till ytterligare förvirring kring begreppet. Förslag på definitioner är enligt Häggström "automatiserat beslutfattande" och "effektiv optimering". Vidare så krävs det ett visst mått av komplexitet för att undvika att system för exempelvis en termostat ska klassas som AI.

Som ett resultat av att en standardiserad nomenklatur inom AI-området saknas har begreppen upplärd AI och självlärd AI introducerats i denna studie, i syfte att hålla isär olika typer av AI. En upplärd AI har enligt rapportens egen definition byggts utan någon form av maskininlärning, vilket i praktiken betyder att all kod är genererad av en människa. Detta skulle exempelvis kunna vara någon form av beslutsträd. Självlärd AI är däremot en modell som byggts genom användning av maskininlärning. Detta illustreras nedan i figur 1.



Figur 1: Träd som skiljer upplärd AI från självlärd AI.

AI är ingen ny teknik, redan 1979 blir backgammonprogrammet BKG första program att besegra en världsmästare i någon form av spel. 1994 fortsätter AI:s framgångar när CHINOOK vinner världsmästerskapet i dam och blir därmed första program att koras till

världsmästare i ett skicklighetsbaserat spel (Bostrom 2014). Bostrom beskriver två olika typer av AI, snäv och generell. BKG och CHINOOK är två exempel på snäv AI. Snäv AI är program som endast klarar av en specifik uppgift, CHINOOK spelar dam och skulle inte klara av att spela schack. Andra exempel på snäv AI som är mer vardagligt förekommande är Googles sökalgoritm och Apples röstassistent Siri. Var gränsen mellan allmän mjukvara och AI går är inte klarlagt och program som klassas som AI av vissa behöver eventuellt inte klassas som det av andra.

Generell intelligens är AI som har en övermänsklig intelligens inom ett område (Bostrom, 2014). Till exempel skulle en ingenjörs-AI utklassa alla mänskliga områdesexperter inom allt som rör ingenjörskonst. Denna generella intelligens är emellertid i dagsläget inte tekniskt möjlig, utan förmodligen några årtionden bort.

3.1.1 Maskininlärning

1965 publicerades Nils Nilssons bok “Learning Machines: foundations of trainable pattern-classifying systems” (Nilsson, 1965). Nilsson ämnade att presentera resultat från det då nya och spännande området inlärningsmaskiner. Dessa maskiner definieras genom att deras handlingar påverkas av dåtida erfarenheter. Idag används istället uttrycket maskininlärning eller maskinell statistisk inlärning. Maskininlärning definieras som ett datorprogram vilket innehar förmågan att lära sig från erfarenheter utifrån hur den tidigare presterat på en viss uppgift (Mitchell, 1997). Till exempel kan ett program designat för att lära sig spela dam förbättras utifrån antalet segrar som prestationsmått, där uppgiften är att spela och erfarenheten är tidigare matcher.

Maskininlärning kombinerar datavetenskap med matematik och statistik för att skapa modeller som kan anpassa sig (Marsland, 2014). Med detta menas att parametrar kan modifieras i en modell så att outputen bättre representerar det förväntade resultatet, givet en viss input. På så sätt kan modellen lära sig och får en generaliserad förståelse för ett visst område, det vill säga att den klarar av att generera rimliga förutsägelser från data som inte fanns med vid inlärningen.

Marsland (2014) nämner två typer av problem där maskininlärning är användbart. Den första typen är så kallade regressionsproblem, vilket är ett problem där en beroende variabel är förutsedd med hjälp av en uppsättning inputvariabler. Ett exempel på ett sådant problem är hur ett elföretag kan använda sig av inputvariabler som historisk förbrukning, väder och säsong samt vad det är på tv för att förutspå hur mycket el som kommer förbrukas vid vissa tidpunkter i framtiden, vilket motsvarar modellens outputvariabel.

Den andra typen är så kallade klassificeringsproblem, vilket är en typ av problem där output är vilken av en mängd klasser varje set av inputvariabler passar in i (Marsland, 2014). Ett exempel på detta är en modell som klassificerar låntagare i olika riskgrupper, där varje grupp har olika risk att inte kunna betala tillbaka. Inputvariabler i ett sådant system skulle kunna vara ålder, total skuld och lön.

Det är också värt att notera att vissa problem kan formuleras om så att de byter typ (Marsland, 2014). I låneproblemet skulle exempelvis regression kunna nyttjas för att direkt föreslå ett lämpligt tak för summan pengar en individ får låna. Maskininlärning har alltså två olika sätt att generera resultat, antingen som ett set av kontinuerliga värden eller som en klassificering av värden. Oavsett typen av problem kan maskininlärning ses som en form av funktionsapproximering, där ett set av inputvariabler används för att beräkna ett rimligt utfall. Eftersom maskininlärning innebär att algoritmer lär sig av data så är både data och hanteringen av denna ytterst viktig. Principen “garbage in, garbage out” är därför mycket relevant inom området, då algoritmer som ges opassande eller motsägelsefulla data kommer resultera i modeller som genererar värdelösa förutsägelser (Geiger et al., 2019).

Data av hög kvalitet är emellertid inte det enda kravet för att bygga en fungerande modell, då misstag kan göras vid förberedelsen av data och inlärningsprocessen i sig (Marsland, 2014). För att undvika att sådana misstag byggs in i modellen finns det några aspekter som är viktiga att tänka på. Det måste först tas ett beslut gällande vilken typ av data som ska samlas in, samt hur stor denna mängd data skall vara. Detta är två faktorer som kommer ha stor påverkan på inlärningen. För att besluta om vilken typ av data som skall samlas in så krävs det att man vet vilka faktorer som påverkar det som skall modelleras. Till exempel korrelerar vädret med hur mycket energi som förbrukas i uppvärmningen av byggnader, medan dagens aktiekurs inte

gör det. Hur stor mängd data som ska samlas in bestäms ofta ur ett kostnadsperspektiv, då det ur modelleringssynpunkt sällan är negativt att samla in mer data.

Utöver att samla in rätt data är det också viktigt att utvärdera modellen med oberoende data. Den data som modellen tränas på kallas för inlärningsdata (Marsland, 2014). Denna data är endast exempel på olika typer av input från det område som modellen förväntas göra förutsägelser på. Inlärningsdata kan å andra sidan omöjligtvis vara en komplett representation av verkligheten, eftersom det då inte funnits ett behov av generaliserad kunskap, vilket innebär att ingen maskininlärning är nödvändig. Om inlärningsdata inte är representativ för den data som modellen kommer stöta på vid användning, kan det resultera i en vinklad modell. Det existerar även ett allvarligare problem, överanpassning. Detta problem kan uppstå vid inlärningsprocessen och grundar sig i att antalet gånger som algoritmen måste granska datasetet och ändra sina parametrar för att en användningsbar modell ska skapas ofta är svårt att bestämma. Utöver detta så är data sällan en ren representation av det underliggande mönstret algoritmen avser att hitta, utan innehåller oftast ytterligare variationer, kallat orelaterat oljud. Att separera underliggande data från oljud är svårt för en inlärningsalgoritm, då den endast ser datapunkter. En algoritm som tränas för mycket kan alltså basera förutsägelser på detta oljud. Enkelt uttryckt kan en modell övertränas, vilket leder till att modellen, på grund av oljudet, producerar output som är orelaterad till input.

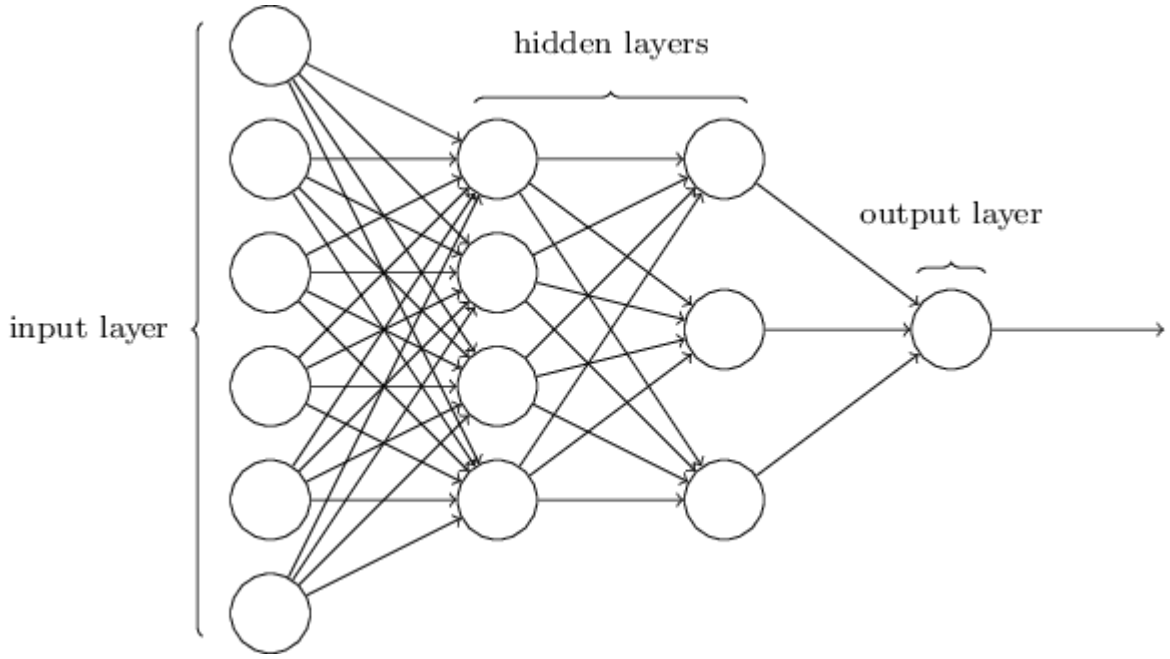
3.1.1.1 Modeller för maskininlärning

Nedan presenteras två olika typer av modeller för maskininlärning. De är på inget sätt representativa för området som helhet utan ämnar endast påvisa att det finns flera metoder för att skapa en självlärd AI, med olika för- och nackdelar.

3.1.1.1.1 Neurala nätverk

Termen neurala nätverk kommer ifrån försök att efterlikna och skapa matematiska representationer för hantering av information i biologiska system (Bishop, 2006). Neurala nätverk är i grund och botten grafer bestående av noder och kanaler mellan dessa (Jordan, 2014). Dessa noder kallar för neuroner. De fungerar genom att ta ett eller flera input värden och använda dessa för att returnera ett värde mellan 0 och 1. Dessa neuroner kan sedan användas för att bygga upp ett neuralt nätverk. Ett neuralt nätverk består av tre lager, input, hidden och output, se figur 2. Input- och outputlagren är relativt självförklarande och att ett

lager är gömt betyder egentligen att det varken innehåller input eller output. Värt att tillägga är att det inte finns någon bestämd mängd neuroner i varje lager (Nielsen, 2019).



Figur 2: Figuren illustrerar ett neuralt nätverks uppbyggnad (Nielsen, 2019).

För att exemplifiera ett neuralt nätverk skapas en modell som kan bestämma om en handskriven siffra är en nia eller inte (Nielsen, 2019). Låt säga att modellen matas med en 64x64 bild i gråskala då skulle inputlagret bestå av 64x64 = 4096 neuroner, en för varje pixel. Output-lagret skulle då endast ha en neuron som returnerar ett värde större än 0,5 om bilden föreställer en nia och mindre än 0,5 om inte. Att designa hidden layer är däremot inte lika enkelt och det saknas enkla regler och riktlinjer för hur det ska gå till.

Ett neuralt nätverk har förmågan att lära sig icke-linjära komplexa förhållanden (Mahanta, 2017). Denna förmåga är ytterst viktig för en AI att inneha då många verkliga relationer mellan input och output är icke-linjära såväl som komplexa. Ytterligare en fördel med neurala nätverk är att de har förmågan att generalisera. När ett neuralt nätverk har lärt sig från de första inputvariablerna och förstått relationerna mellan dessa så kan den antyda samband mellan osedda data. Detta tillåter modellen att generalisera och generera förutsägelser på tidigare osedda data. Till skillnad från många andra metoder för förutsägelse så ställer ett neuralt nätverk inga krav på inputvariablerna, till exempel hanterar de data med hög volatilitet och icke-konstant varians bättre än andra modeller.

En av de största nackdelarna med neurala nätverk är att modellen ofta tenderar att bli en svart låda utan insyn för användaren (Mahanta, 2017). Utvecklaren av modellen ställer upp den data som nätverket ska tränas på och låter sedan den själv bestämma vilka variabler som är viktigast. Det är därmed svårt att redogöra för vilka variabler som väger tyngst i en viss output. En annan nackdel med neurala nätverk är att de löper stor risk att övertränas. Detta kan dock förhindras genom att bättre hantera och validera data och modell.

Neurala nätverk har idag flertalet användningsområden, däribland analys av tal, objektigenkänning, ansiktsigenkänning och analys av handskriven text (Mahanta, 2017). Neurala nätverk kan även användas till dataklassificering och mönsterigenkänning.

3.1.1.1.2 Bayesiska nätverk

Ytterligare en modell för maskininlärning är grafiska modeller, mer kända som bayesiska nätverk. Dessa nätverk är så kallade riktade acykliska grafer (directed acyclic graphs eller DAGs). Varje nod i detta nätverk representerar de variabler som är intressanta, och länkarna mellan noderna representerar den kausala påverkan, eller det orsaksförhållande, en variabel har på en annan (Pearl, 2014).

Skillnaden mellan de flesta andra modellers sätt att representera kunskap och bayesiska nätverk, är att bayesiska nätverk modellerar miljön snarare än resoneringsprocessen (Pearl 2014). Faktumet att det är kausala mekanismer i miljön som modelleras gör att det är möjligt att ställa olika typer av frågor, såsom associativa frågor: “Givet A vad kan vi förvänta oss av B?”, abduktiva frågor: “Vilken är den mest rimliga förklaringen för ett givet set av observationer?” och kontrollfrågor: “Vad händer om vi agerar på omgivningen?”. Associativa frågor besvaras med probabilistisk kunskap om modellen, medan de två andra frågornas svar bygger på den kausala kunskapen som existerar inom nätverket. Både associativ och kausal information kan effektivt representeras i ett bayesiskt nätverk.

Den associativa aspekten hos bayesiska nätverk gör att de lämpar sig för kognitiva uppgifter, såsom objektigenkänning, läsförståelse och temporal projektion (Pearl, 2014). Temporal projektion innebär att med utgångspunkt i ett initialt tillstånd och en eller flera händelser som inträffar så resoneras utifrån tillståndet som följer av händelserna. Ett exempel på detta är att

Lisa går in i köket, vilket följs av att diskhon svämmar över och att katten äter all mat. Att analysera sambanden mellan dessa händelser involverar temporal projektion (Mueller, 2015).

Bayesiska nätverks mest distinkta egenskap är deras förmåga att representera och agera på ändrade konfigurationer (Pearl, 2014). Det är relativt enkelt att ändra på ett nätverks topologi och denna flexibilitet tillåter modellen att hantera olika situationer utan att tränas om.

3.2 Existerande juridiska tillämpningar av AI

Det finns redan idag ett flertal exempel på hur AI har implementerats, både inom domstolsväsendet och andra myndighetsfunktioner. I detta kapitel kommer några av systemen att beskrivas och deras respektive mottagande i samhället att presenteras.

3.2.1 Trelleborgsmodellen

I Trelleborgs kommun är ansökningsprocessen för försörjningsstöd sedan 2015 digitaliserad (Svensson & Larsson, 2017). Enligt kommunen har denna modell effektiviserat hanteringen av ärenden avsevärd. Innan implementeringen kunde handläggningstiden uppgå till 20 dagar, medan ett ärende i dagsläget, i bästa fall, kan avklaras på under ett dygn.

Trelleborgsmodellen använder sig av RPA-algoritmer (Svensson, 2019), dessa ingår i vad som i rapporten benämns som upplärda algoritmer. RPA arbetar utifrån en rad av fördefinierade regler och algoritmer, vilket är fördelaktigt när ett enkelt, repetitivt arbete ska utföras. I Trelleborgsmodellens fall tar roboten alltså bara beslut utefter de regler som har satts upp. På samma sätt kan inte systemet hantera ett undantag från de regler som har satts upp. I ett sådant fall måste beslutet tas av en människa.

Det är emellertid inte bara i Trelleborg som denna modell för försörjningsstöd används. Enligt Knutas & Bergström (2018) bedriver 14 kommuner sedan 2017 ett projekt där de jobbar för att ta fram varianter av Trelleborgsmodellen till sina egna verksamheter. Detta har dock mötts med starkt motstånd från socialarbetare i vissa av de berörda kommunerna. I Kungsbacka kommun, som är en av de kommunerna där en modell liknande den i Trelleborg utvecklades, sade 12 av de 16 handläggare upp sig som ett resultat av kommunens tilltänkta implementering. Missnöjet med att automatisera vissa av socialtjänstens funktioner återfinns

även i en studie av Gabriella Scaramuzzino (2019). Här har medlemmar i Akademikerförbundet SSR tillfrågats om deras åsikter kring automatisering av socialt arbete. Bland de tillfrågade svarar 86 procent att de instämmer mycket eller ganska mycket på påståendet "En robot eller programvara kan inte skapa sig en helhetssyn på en individs eller en familjs situation". I samma studie svarar 81,9 procent av de tillfrågade att de instämmer mycket eller ganska mycket på påståendet "Automatisering hindrar att sociala problem fångas upp (t.ex. missbruk och barn som far illa)".

3.2.2 Brottsförutspående programvaror

Det finns även algoritmer vars syfte är att förutspå framtida brottslighet, där två av de mest kända inom området är COMPAS och SyRI. COMPAS, eller Correctional Offender Management Profiling for Alternative Sanctions, är ett system som skapades av mjukvaruföretaget Northpointe redan år 1998 (MacMillen, 2019). Syftet med COMPAS är att ge en bedömning av risken för att en frigiven brottsling inom två år kommer att begå ytterligare brott. COMPAS tar vid en bedömning 137 olika faktorer i beaktning, däribland den aktuella personens tidigare brottsregister. COMPAS används till exempel under förundersökningar eller vid beslut om villkorlig frigivning. De som förespråkar detta system menar att det kan göra analyser kring återfallsrisker mer korrekta och mer opartiska än vad en människa skulle kunna göra. Detta är emellertid omtvistat. I en studie av Julia Dressel och Hani Farid (2018) ifrågasätts hur bra bedömningarna som COMPAS utför egentligen är. I studien fick 462 slumpvis valda testpersoner ta del av information om ett antal förbrytare, däribland dess kön, ålder och tidigare brottsregister. Testpersonerna fick dock inte reda på förbrytarens etnicitet. Dessa testpersoner fick sedan avgöra om de trodde att den förbrytare som analyserades hade begått brott igen inom två år från den tidpunkt som de blev frisläppta. Testpersonerna lyckades i 67 procent i fallen svara rätt på om huruvida förbrytaren hade återfallit i brott eller inte. Motsvarande siffra för COMPAS för samma förbrytare var 65 procent.

Ett annat problem som har framhävts med COMPAS är dess etniska partiskhet. I en artikel av Julia Angwin et.al. (2016) visas på hur COMPAS gör helt olika riskbedömningar beroende på vilken etnicitet förbrytaren har. Exempelvis jämförs en mörkhyad 18-årig flicka, som tillsammans med en kompis hade stulit en cykel, med en vit 41-årig man som hade stulit verktyg till ett värde av 86 dollar från en närliggande byggvaruhandel. Även om flickan var

rapporterad för förseelser som minderårig så var det mannen som hade det grövre brottsregistret av de båda, då han tidigare bland annat hade avtjänat ett femårigt fängelsestraff för väpnat rån och försök till väpnat rån. Trots skillnaderna i deras tidigare brottsregister så var det flickan som fick den högsta återfallsrisken, en åtta på en tiogradig skala, vilket kan jämföras med mannen, som blev klassad som en trea på samma skala. Två år efter att bedömningen utförts på dem båda, hade flickan inte blivit åtalad för några fler brott, medan mannen avtjänade ett åttaårigt fängelsestraff efter att ha utfört en grov stöld mot en elektronikbutik.

Enligt Angwin et al. (2016) är detta heller inte det enda fallet där COMPAS har gjort olika bedömningar utifrån människors etnicitet, vilka senare har visat sig vara felaktiga. De menar att i 44,9 procent av de fall där en afroamerikan har bedömts ha en hög risk har inget återfall skett. Bland vita amerikaner med samma riskklassificering var motsvarande siffra 23,5 procent. Dessutom återfaller endast 28 procent av de afroamerikaner som har bedömts ha en låg risk för återfall, medan det bland vita amerikaner med samma riskklassificering var 47,7 procent som begick ytterligare brott.

Det finns emellertid kritik mot artikeln publicerad i ProPublica. Northpointe, företaget bakom COMPAS, hävdar att det finns signifikanta fel i artikeln (Dietrich, Mendoza & Brennan, 2016). Northpointe anser att det i artikeln har fokuserats på statistik som inte har tagit hänsyn till grundnivåerna för återfall bland afroamerikaner och vita amerikaner. De påstår vidare att när korrekt statistik har använts finns det inga tecken på att COMPAS gör några felaktiga bedömningar utifrån människors etnicitet. Northpointe får uppbackning av tankesmedjan Community Resources for Justice, vilka i en rapport riktar kritik mot artikeln i ProPublica (Flores, Lowenkamp & Bechtel, 2016). I rapporten skriver författarna att analysen i ProPublica-artikeln inte följer de accepterade standarder som finns för att testa partiskhet i riskutvärderingar, samt att artikeln endast inkluderade åtalade som ännu inte hade blivit prövade i en rättegång, trots att detta inte var syftet med COMPAS.

I Nederländerna fanns fram till nyligen ett system som liknade COMPAS. Systemet gick under namnet System Risk Indication, eller SyRI, och var ett system som användes för att hitta personer som ansågs ligga i riskzonen för att begå bidragsbedrägeri (Osborne, 2020). SyRI sammanställde typiska egenskaper från personer som tidigare hade begått brott av

denna typ. SyRI använde sedan denna uppsättning av egenskaper för att leta efter andra individer som potentiellt skulle kunnat begå liknande brott. SyRI avskaffades dock år 2020, efter att en nederländsk domstol kommit fram till att systemet stred mot EU:s regler kring personlig integritet (Dutchnews.nl, 2020). Det finns även kritiker som menar att SyRI ledde till diskriminering. En av dessa är FN:s expert inom mänskliga rättigheter Philip Alston, som menar att SyRI kan diskriminera de allra fattigaste (United Nations Humans Rights, 2019). Han hävdar att en bidragande orsak till detta är den bristande transparensen i systemet, samt en avsaknad av debatt och nyhetsbevakning innan introduktionen av systemet.

3.3 AI och etik

Användning av AI i offentlig sektor ställer höga krav på systematisk hantering av etiska överväganden (Myndigheten för digital förvaltning, 2020). I avsnittet nedan presenteras olika riktlinjer som tagits fram av råd och kommissioner, med fokus på etisk implementering och användning av AI.

3.3.1 Algoritmisk rättvisa och bias

Artificiell intelligens har redan implementerats inom flertalet olika områden, däribland inom områden som kan påverka en människas liv i stor grad, exempelvis COMPAS och Trelleborgsmodellen. Det har bidragit till ett ökat fokus på områden som rättvisa och bias inom AI. Det finns ett flertal olika definitioner av rättvisa, varav ett antal kommer att presenteras i kommande avsnitt.

Corbett-Davies, Pierson, Feller, Goel och Huq (2017) nämner tre populära definitioner av algoritmisk rättvisa, där grupperna syftar till den gruppindelning som blir utifrån känsliga attribut:

- Statistisk paritet: En lika stor andel ur grupperna får lika utfall.
- Villkorad statistisk paritet: Om ett eller flera villkor är uppfyllda så får en lika stor andel ur grupperna lika utfall.
- Prediktiv jämlikhet: Träffsäkerheten i utfallet är lika i de båda grupperna, mätt genom andelen falsk positiv.

Som exempel kan COMPAS användas för att besluta om en misstänkt ska bli kvarhållen eller frisläppt i väntan på rättegång. Används exempelvis etnicitet som gruppindelning så innebär statistisk paritet att en lika stor andel av de ljushyade som av de mörkhyade hålls kvar. Villkorad statistisk paritet innebär att av alla som har samma antal fällande domar sedan tidigare så hålls en lika stor andel ur båda grupper kvar. I fallet med prediktiv jämlikhet är det istället andelen av de som inte hade begått ett grovt brott om de hade frisläppts, men som ändå kvarhölls, som ska vara lika.

Corbett-Davies och Goel (2018) nämner tre liknande definitioner för rättvisa:

- Anti-klassificering: Vid beslut tas inte hänsyn till de känsliga attributen och inte heller till de attribut som kan användas för att identifiera grupper utan att känsliga attribut används.
- Klassificeringsparitet: Liknande prestanda i de olika grupperna, mätt i exempelvis falska positiva.
- Kalibrering: Utfallet är oberoende av känsliga attribut, lika stor andel ur båda grupperna utifrån exempelvis riskpoäng.

De skriver vidare att dessa definitioner lider av begränsningar och att de i vissa fall har en negativ inverkan på rättvisan. Som exempel nämner de att kvinnor har en lägre återfallsrisk än män och om inte hänsyn tas till kön i riskbedömningen så kan det leda till hårdare åtgärder mot kvinnor. Klassificeringsparitet kan leda till att undergrupper klassificeras felaktigt för att uppnå liknande prestanda inom grupperna.

Även EU:s expertgrupp för artificiell intelligens (2019) nämner att det finns många definitioner av rättvisa och att de anser att det finns både en materiell och en processuell dimension. Den materiella aspekten berör både fördelningen av fördelar och kostnader, samt tillgång till utbildning, samhällsservice, teknik, etcetera. Grupper och individer ska inte heller vara utsatta för diskriminering och stigmatisering. Den processuella dimensionen syftar till möjligheten att kunna utmana beslut tagna av AI och de personer som hanterar AI:n och för att kunna göra det är det viktigt att kunna identifiera de ansvariga samt ha en hög förklarbarhet.

Bias är ett begrepp som det diskuteras mycket kring när det kommer till rättvisa. Begreppet saknar dock en bra svensk översättning, därav används det engelska ordet genomgående i rapporten. Bias innebär enligt Cambridge Dictionary att något eller någon motsätter sig eller stödjer en särskild person eller sak på ett orättvist sätt, på grund av personliga åsikter som påverkar omdömet (Bias, u.å.).

Det finns ett antal exempel på fall där bias, utan intention från utvecklare, har givit oönskade resultat. Ett exempel är fallet där Googles algoritm för bildigenkänning klassificerade mörkhyade personer som gorillor och där lösningen blev att de helt enkelt tog bort gorilla ur algoritmen (Wired, 2018). Ett annat exempel är det tidigare nämnda COMPAS, där det visat sig att mörkhyade personer bedöms löpa högre risk för återfall än ljushyade personer (Angwin et al., 2016).

Baer (2019) beskriver hur bias kan uppstå på flera olika sätt och hur dessa sätt kan vara mer eller mindre svårhanterliga. Enligt honom är det mest svårhanterliga fallet det fall då algoritmen avspeglar samhällsbias, då det kan framstå som korrekt i samhällets ögon. Här kan bias i mänskligt beslutstagande vara en stor faktor och som exempel nämns hur polisens agerande kan påverka data som används för att ta fram en algoritm för att identifiera droginnehav. Har polisen en tro att en grupp oftare har droger på sig så riskerar det att leda till att de är mer noggranna vid visitationer och att de på så sätt oftare finner droger. När sedan den data används för att skapa algoritmen kan det leda till att attribut kopplade till den gruppen får en ökad vikt.

Vidare skriver Baer (2019) om bias introducerad av utvecklaren av algoritmen, här nämns tre olika varianter:

- Bekräftelse-bias: Utvecklarens egen bias reflekteras i modellen, detta kan påverka både skapandet av modellen och insamlandet av data.
- Ego-utarmning: Utvecklaren missar möjligheten att undvika bias på grund av distraktion/mental trötthet.

- Överdrivet självförtroende: Utvecklaren ignorerar indikationer på att modellen kan innehålla bias.

När det kommer till bias i data som används för algoritmen så kan det införas bias vid olika steg i dataprocessen:

- Datakällan kan bidra till bias.
- Bias kan uppstå vid användning av data i utveckling av modeller.

Enligt Baer (2019) finns även en risk för stabilitetsbias för en algoritm, vilket syftar till att en algoritm är framtagen med hjälp av historiska data och det kan därför leda till att algoritmen inte längre är relevant om omvärlden förändras strukturellt. Stabilitetsbias kan även uppstå om data som använts inte har samlats in under tillräckligt lång tid och inte täcker en hel cykel till exempel.

Slutligen skriver Baer (2019) om bias som algoritmen själv introducerar. Det kan exempelvis bero på låg frekvens av ovanliga tillstånd och som därmed slås ihop med andra grupper och därmed kan beteenden från dessa grupper avspegla sig på det ovanliga tillståndet.

3.3.2 Befintliga riktlinjer för AI

European Commission for the Efficiency of Justice, CEPEJ (2018), skriver i sina stadgar om fem huvudsakliga principer för användning av AI inom det juridiska väsendet. Den första berör respekt för fundamentala rättigheter, och avser att bearbetning av data och beslut måste ha ett tydligt syfte, samt följa grundläggande mänskliga rättigheter. Utöver detta måste också säkerställas att allas möjlighet till en rättsprocess, som dessutom ska vara rättvis, inte undermineras då ett AI-verktyg stödjer eller självt tar juridiska beslut. Att föredra är därför verktyg där designen förhindrar att de mänskliga rättigheterna och andra etiska grundprinciper bryts.

CEPEJ:s (2018) andra princip finns för att hindra förekomst och/eller intensifiering av diskriminering, exempelvis bearbetning av uppgifter som kan vara av känslig natur, vilka inkluderar bland annat information om etnisk och socioekonomisk bakgrund, kön och

politiska åsikter. Den tredje behandlar bearbetningsprocessens kvalitet och säkerhet, vilket innefattar att algoritmer lagras och körs i miljöer som garanterar att systemet har önskad integritet och abstraktion. Data som hämtats från historiska juridiska beslut bör endast komma från certifierade källor och ska inte modifieras innan dess att de har använts av inlärningsmaskiner. Nästa princip rör opartiskhet, transparens och rättvisa, att säkerställa en godtagbar grad av förståelse och tillgänglighet. Detta kan ske genom användande ett tekniskt sett helt transparent system, som till exempel använder öppen källkod, men finns det risk att känslig information hamnar i oönskade händer. Alternativet är ett system där resultaten förklaras tydligt med allmän information om hur tjänsterna och verktygen fungerar, samt möjliga felorsaker. Den sista principen åsyftar att användare ska vara informerade nog för att vara i kontroll över plattformen, att användarens autonomi ökas snarare än begränsas när AI:n nyttjas. Det ska finnas tydlig information tillgänglig om huruvida ett domstolsfall har processats av en AI innan eller under en juridisk process, och användaren ska kunna invända mot AI:ns utfall.

Utöver dessa finns etiska riktlinjer för mer generell användning av AI framtagna av EU:s High Level Expert Group on Artificial Intelligence (2019). Där nämns 3 egenskaper som gör en AI trovärdig, den ska följa alla lagar som kan appliceras, den ska följa befintliga etiska principer och värderingar och den ska vara tekniskt och socialt robust för att undvika förekomst av oavsiktlig skada. Var och en av dessa tre egenskaper är nödvändig, och för att en AI ska betraktas som trovärdig behöver samtliga finnas.

I syfte att säkerställa förekomsten av dessa egenskaper har AIHLEG (2019) i sina riktlinjer också tagit fram sju krav som ställs på AI som utvecklas.

- Mänskligt agentskap och mänsklig tillsyn: Systemet som implementeras bör stödja människans autonomi och beslutsfattande. Då krävs att det möjliggör ett demokratiskt och jämlikt samhälle genom att stödja mänskliga rättigheter.
- Teknisk robusthet och säkerhet: AI-system måste vara säkra och motståndskraftiga. De måste innehålla en reservplan ifall något går fel, samt vara pålitlig, reproducerbar och noggrann. Detta kan potentiellt minimera och förebygga oavsiktlig skada.

- Integritet och dataförvaltning: Ett AI-system måste genom hela sin livscykel garantera att känsliga uppgifter skyddas och att integritet säkerställs. Dessutom får de dataset som systemet tränas på inte innehålla någon oönskad bias.
- Transparens: Vid en implementering av en AI måste det finnas transparens i både data, systemet och affärsmodellerna. Det är också viktigt att det finns en förklarbarhet i AI-systemet.
- Mångfald, icke-diskriminering och rättvisa: Det är viktigt att möjliggöra inkludering och mångfald under AI-systemets hela livscykel och lika tillträde genom inkluderande utformningsprocesser och likabehandling.
- Samhällets och miljöns välbefinnande: Hållbarhet och ekologiskt ansvarstagande i AI-system och forskning om AI-lösningar som rör globala frågor skall uppmuntras. AI bör användas till nytta för samtliga människor, både nuvarande och kommande generationer.
- Ansvarsskyldighet: Det behövs mekanismer för att säkerställa ansvarstagande och ansvarsskyldighet för AI-system och deras resultat.

3.4 FN:s globala hållbarhetsmål

Ur ett etik- och hållbarhetsperspektiv har ett antal av FN:s (2015) 17 globala hållbarhetsmål identifierats, vilka i någon mån kan appliceras på studien.

Dessa mål är:

- 5: Jämställdhet
- 8: Anständiga arbetsvillkor och ekonomisk tillväxt
- 10: Minskad ojämlikhet
- 16: Fredliga och inkluderande samhällen
- 17: Genomförande och globalt partnerskap

Mål nummer fem berör jämställdhet mellan kvinnor och män. FN (2015) definierar jämställdhet som en rättvis fördelning av makt, inflytande och resurser. Att leva ett liv fritt från våld och diskriminering är enligt FN en grundläggande rättighet som är nödvändig för att ett samhälle ska nå sin fulla potential.

Anständiga arbetsvillkor och ekonomisk tillväxt är FN:s (2015) åttonde globala mål och verkar för att uppnå en varaktig, inkluderande och hållbar ekonomiskt tillväxt samt anständiga arbetsvillkor för en produktiv sysselsättning. Delmål 8.3, främja ekonomisk produktivitet genom diversifiering teknisk innovation och uppgradering, uppnås genom att med hjälp av bland annat ny teknik uppnå en mer produktiv ekonomi.

FN:s (2015) tionde mål, minskad ojämlikhet bygger principen om allas lika rättigheter oberoende av faktorer såsom kön, etnicitet och religion. Att öka jämlikheten i ett samhälle minskar risken för konflikter och främjar människors förmåga att delta i samhällsutvecklingen.

Det 16:e hållbarhetsmålet som FN (2015) har tagit fram, fredliga och inkluderande samhällen innefattar bland annat att alla människor är lika inför lagen och ska ha samma tillgång till rättvisa. Att främja rättssäkerhet och tillgången till rättvisa är delmål 16.3 och syftar på att förbättra dessa två faktorer både nationellt och internationellt.

Det sista målet, nummer 17, heter genomförande och globalt partnerskap. FN (2015) menar att internationella investeringar och samordnad politik krävs för en nyskapande teknisk utveckling. Att utbyten av kunskap, expertis, teknik samt finansiella resurser sker är nödvändigt för att målet ska uppnås. Delmål 17.6 berör att samarbeta och dela kunskap kring vetenskap, teknik och innovation syftar att säkerställa kunskapsutbyten mellan länder.

4. Resultat och analys

I detta kapitel presenteras resultatet från intervjustudien. Respondenternas svar och den diskussion som förts under intervjuerna har delats in teman för att få en tydlig struktur och bra överblick. Dessutom har respondenterna subjektivt delats in i någon av kategorierna juridik eller AI utifrån deras utbildning, yrkesbakgrund och nuvarande befattning. I tabell 3 presenteras kodningen av studiens teman.

Tabell 3: Presentation av avsnittets uppdelning, fördelat på två nivåer.

Huvudtema	**Undertema**
Positiva implikationer	Kostnader och effektiviseringar
	Domstolarnas kvalitet och enhetlighet
Risker och begränsningar	Transparens och förklarbarhet
	Bias och representativa data
	Kunskap och kompetens
	Tillit och legitimitet
	Tröghet inom det juridiska väsendet
	Kritik mot befintliga system
Implementeringsaspekter	Användningsområden
	Placering och fördelning av ansvarsbördan
	Prestandakrav på AI
	Möjlighet till överklagan
	Metodik för implementering och lämpliga metoder

Respondenternas inställning	

4.1 Positiva implikationer

I denna kategori presenteras de teman som lyfter de positiva implikationerna av att implementera AI i svenska domstolar.

Tabell 4: I tabellen presenteras exempelcitat från området Positiva implikationer.

Tema	**Exempelcitat**
Kostnader och effektiviseringar	*"Det finns ju mycket uppgifter som vi gör som är kanske lite mer rutinartade som handlar om att söka igenom stora mängder information, få fram rätt information och strukturera det på ett bra sätt. Det kan ju en maskin göra betydligt snabbare."* - Johan Sanner
Domstolarnas kvalitet och enhetlighet	*"Vid användning är de stora möjligheterna såklart att man kan få bättre och mer korrekta och likriktade domar."* - Karl de Fine Licht

4.1.1 Kostnader och effektiviseringar

I följande tema redogörs för vilka möjligheter respondenterna ser i form av minskade kostnader och effektivisering med hjälp av AI.

4.1.1.1 Juridik

En majoritet av juridikrespondenterna menar att effektivisering och kostnadsbesparing är två av de största fördelarna med att implementera AI i Sveriges domstolar. Detta kan exempelvis ske genom en delvis automatiserad domstol där monotona arbetsuppgifter sköts av en dator.

Exakt vilka dessa arbetsuppgifter skulle kunna vara saknar konsensus, men ett par relativt okontroversiella exempel är arkivering och datainsamling.

"Det finns ju mycket uppgifter som vi gör som är kanske lite mer rutinartade, som handlar om att söka igenom stora mängder information, få fram rätt information och strukturera det på ett bra sätt. Det kan ju en maskin göra betydligt snabbare."

- Johan Sanner

Att en AI kan hantera stora mängder data och bearbeta den snabbare än en människa är ytterligare en fördel som några respondenter har lyft. En dator behöver heller inte vila utan kan bearbeta data dygnet runt.

"De uppenbara är väl att AI inte behöver förhålla sig till arbetstider, kan processa stora mängder data."

- Markus Naarttijärvi

4.1.1.2 AI

Bland AI-respondenterna är en av de mest frekvent nämnda fördelarna att det går att göra stora besparingar i den offentliga sektorn om enklare processer automatiseras. I intervjuerna belyses att dagens arbetssätt i den offentliga sektorn är långsiktigt ohållbart och att effektiviseringen som implementeringen av AI skulle innebära i framtiden är ett måste. Denna automatisering leder till att resurser istället kan flyttas till mer komplexa och krävande arbetsuppgifter.

"DIGG, Digitaliseringsmyndigheten, de gjorde en rapport precis innan jul där de kom fram till att offentlig sektor skulle kunna spara 140 miljarder på att använda AI i offentlig sektor och det var ändå bara 'vad är det man gör idag som skulle kunna, ja, automatiseras eller underlättas på olika sätt'. Så att självklart ser jag extremt stora möjligheter. [...] *Offentlig sektor har inget rent val, eller snarare om man inte drar nytta av den här tekniken så kommer det att bli ännu värre..."*

- Fredrik Heintz

4.1.2 Domstolarnas kvalitet och enhetlighet

I detta tema lyfts respondenternas åsikter kring AI:s möjligheter att förbättra de svenska domstolarnas kvalitet och enhetlighet.

4.1.2.1 Juridik

Flertalet av juridikrespondenterna nämner att AI kan bidra till ett mer enhetligt dömande och en höjd kvalité inom domstolarna. En AI skulle dessutom kunna bidra med flera olika typer av stöd i en rättssal, exempelvis i form av dokumentgranskning eller översättning. Detta skulle i sin tur kunna leda till att Sveriges domstolar får ett högre förtroende. Ytterligare en fördel, lyft av John Lagström, är att domstolar skulle med hjälp av AI kunna bli mer tillgängliga och att rättssäkerheten på så sätt skulle förbättras.

> *"De stora fördelarna som jag ser är att vi kan höja kvaliteten inom olika delar av rättsväsendet, avsevärt. Till exempel när det kommer till tolkning så har vi möjligheten att säkerställa att tolkningen alltid blir rätt och att den görs på ett jämlikt och jämställt sätt*."
>
> - John Lagström

Det finns även respondenter som nämner att en AI kan användas för att kontrollera att domstolen prövat alla de sakfrågor som bör prövas.

> *"Då kan det vara bra att veta, har jag prövat A, B, C och D? För ibland kanske man bara har prövat A, B och C. Att man då får 'Nu har du glömt D och har du prövat A, B och C så måste du pröva underfrågorna…'."*
>
> - Anders Hagsgård

Magnus Ivarsson menar å andra sidan att ett enhetlig dömande i sig inte nödvändigtvis är ett helt lämpligt kvalitetsmått samt att enhetlighet inte är uppnåbart om domarnas individuella frihet gällande bedömning ska kvarstå.

> *"Varje domstol är en egen myndighet, och ska man hårdra det så är ju varje domare en egen myndighet.* [...] *Det går inte att förvänta sig en helt jämlik bedömning någonsin, eftersom alla individer som kommer till domstolen har egna unika omständigheter, och domarna som*

dömer är unika domare. Så jag tycker att det är svårt att säga att bara därför att vissa domstolar har dömt ut mer fängelse under en period, så måste man hitta en högre standardisering."

- Magnus Ivarsson

4.1.2.2 AI

Även under intervjuerna med AI-respondenterna har det lyfts fram att det är möjligt att få ett mer konsekvent och rättvist dömande genom att implementera AI inom domstolsväsendet. Med hjälp av en AI finns det potential att få en högre kvalitet i domarna, i och med att domarna underbyggs bättre.

"Vid användning är de stora möjligheterna såklart att man kan få bättre och mer korrekta och likriktade domar."

- Karl de Fine Licht

"*Fördelar: Effektivare och mer konsekventa beslut."*

- Jacob Dexe

4.2 Risker och begränsningar

I följande kategori lyfts de teman som påvisar de risker och begränsningar som följer av att introducera och nyttja AI inom det svenska domstolsväsendet.

Tabell 5: I tabellen redovisas exempelcitat från området Risker och begränsningar.

Tema	**Exempelcitat**
Transparens och förklarbarhet	*"Being too transparent actually doesn't help. If I give you ten thousand lines of code and say, 'this is the program', it would not really help you. You want something that is suited for your background and for your purpose."* - Frank Dignum

Bias och representativa data	*"Det kommer finnas spår av konstiga mänskliga fördomar, begränsningar, kognitiva förmågor i datat, som vi inte kan sätta fingret på när vi tittar på datat, men som kommer spela ut när en AI identifierar trender i datat. Sådana spår kommer finnas med där på något sätt."* - Jacob Dexe
Kunskap och kompetens	*"...Man är fortfarande lite rädd och man är fortfarande lite skeptisk och har inte alltid kompetensen som kanske krävs, för att kunna använda den till fullo då."* - Josefin Rosén
Tillit och legitimitet	*"Idag har ju den enskilda domaren ett högt förtroende i kraft av erfarenhet, utbildning och sin roll. Det är ett etos som är svårslaget. Det blir nog svårt, i alla fall i dagsläget, att låta en AI ta den rollen."* - Magnus Ivarsson
Tröghet inom det juridiska väsendet	*"Juridiken ska vara lite trögrörlig, det ska inte gå för fort, för konsekvenserna om det blir fel är enorma. Ny lagstiftning ska ta tid och förändringar i sättet vi arbetar på måste ta tid, därför att det är en förutsägbarhetsfråga att folk inte gör på helt olika sätt. Då drabbas liksom hela rättsstaten av att det blir spretigt och konstigt och knepigt, och då funkar inte, då kommer vi tappa allt vad gäller legitimitet för juridiken… Sverige är ju ett väldigt modernitetslängtande land. Vi har en strävan efter att vara moderna och vi bygger liksom alltid vårt samhälle på framtid. Vi är inte rädda för det nya."* - Anna-Sara Lind
Kritik mot befintliga system	*"När man använder COMPAS till exempel har man en tendens att förstärka den första magkänslan som man hade kring en person. Om du är negativt inställd till någon som kommer in, och så säger vi att den (COMPAS) ger en sjua av tio, då tenderar man att bli ännu mer säker på sin sak."* - Karl de Fine Licht

4.2.1 Transparens och förklarbarhet

I temat lyfts respondenternas inställning kring transparens och förklarbarhet gällande en eventuell implementering av AI inom Sveriges domstolar.

4.2.1.1 Juridik

En aspekt av transparens och förklarbarhet som framkommit under studien är bristen på definition av begreppet transparens. Är att lämna ut källkod och dataset att vara transparent om det saknas kunskap att tolka och förstå dessa?

Samtliga tillfrågade respondenter anser att transparensproblematiken är viktig att diskutera. Ett antal av respondenterna ansåg att det krävs full transparens för att kunna implementera en AI, vilket innebär att både källkod och dataset är offentligt tillgängliga, i enlighet med offentlighetsprincipen.

"Offentlighetsprincipen bör ju tillämpas på vår beslutsmekanism. Offentlighetsprincipen gäller förhandlingar till exempel, så då tycker jag samma princip ska gälla för hur besluten fattas av en AI."

- Magnus Ivarsson

Vissa respondenter menar att om förklaringen är av tillräckligt hög kvalitet så är det möjligt att källkoden inte behöver vara offentlig.

"Den algoritmen skulle du kunna skriva ut i domskäl och förklara att eftersom detta och detta och detta så blir det här straffet. Då får du en transparens i algoritmen genom att man skriver skälen, kanske man inte behöver lämna ut själva algoritmen i sig."

- Björn Östman

I fall av en självlärd algoritm så menar Markus Naarttijärvi att det dataset som ligger till grund för inlärningen också måste vara tillgängligt för att uppnå full transparens.

"Jag skulle säga att det beror på hur man applicerar de här [AI] *och graden av konsekvenser för den enskilde. Om man tänker sig ett system inom rättsväsendet som skulle kunna komma med domar i mål som rör grundläggande rättigheter på ett eller annat sätt. Då skulle jag säga att kraven på transparens är ganska högt ställda. Även om man skulle säga att den är på en övergripande nivå så drar den slutsatser baserat på kriterium A, B och C, så problemet är att relationen mellan de kriterierna och den data som poängsättningen av respektive kriterium bygger på är väldigt svår att illustrera. Även om den skulle säga att algoritmen har tilldelat dig de här poängen på de här olika kriterierna baserat på en statistisk analys av tidigare fall. I och med att det här är ett maskinlärande system kommer algoritmen att utvecklas för varje datapunkt man matar den med. Transparensen i systemet är beroende av all data man matat den med. En fullständig transparens, vilket vi i och för sig inte i fråga om mänskligt beslutfattande heller har, kräver då tillgång till datasetet."*

- Markus Naarttijärvi

John Lagström lyfter att det är möjligt att kraven på graden av transparens eventuellt kan variera i olika situationer. Det kan finnas situationer som kräver modeller med en hög grad av förklarbarhet samtidigt som det i andra fall duger med en lägre grad. Han är å andra sidan också mycket tydlig med att om Sveriges domstolar ska använda AI så ska det ske på ett fullständigt transparent sätt.

"Det kan mycket väl vara så att viss information ur en datamängd där måste man ha en jättetydlig förklaring om varför det blev som det blev, annan information kanske man kan använda rekursiva neurala nätverk som är helt wide open. Jag tror inte det, men det är möjligt."

- John Lagström

4.2.1.2 AI

Majoriteten av de i studien tillfrågade anser att tydliga förklaringar av algoritmens logik, samt utförliga motiveringar till de resultat som ges bör prioriteras högre ur ett transparensperspektiv än att den bakomliggande koden finns tillgänglig för alla.

"Being too transparent actually doesn't help. If I give you ten thousand line of codes and say, 'this is the program', it would not really help you. You want something that is suited for your background and for your purpose."

- Frank Dignum

Frank Dignum jämför även mot hur det inom domstolarna också är centralt med en förklaring av ett beslut.

"And that's actually very similar to how the court works as well, because the judge doesn't just give a judgement, they also give an explanation of how they've come to that judgement."

- Frank Dignum

Fredrik Johansson menar också att koden inte är det viktigaste för att förklara algoritmens logik. Han konkretiserar också hur en självlärd algoritm ska kunna förklaras.

"Koden säger ju väldigt lite egentligen. Hela poängen med maskininlärning är ju att lära sig från data och då är ju datamängden absolut minst lika viktig. Sen kan man ju se själva outputen från ett maskininlärningssystem som kod om man vill. Till exempel, om systemet lär sig ett beslutsträd eller en annan regelbaserad modell som säger 'om du uppfyller de här villkoren så kommer vi förutspå att du ska begå brott igen, och därför kommer vi göra xx'. I det fallet är outputen, eller 'koden', transparent själv."

- Fredrik Johansson

Flera av respondenterna understryker också faktumet att människan inte heller är särskilt transparent och inte alltid kan motivera sina egna beslut och handlingar.

"Det är ju inte så att vi människor kan förklara alla beslut. Om du pratar med en läkare som tittar på en EKG-kurva och blixtsnabbt säger vad det är, och sedan frågar 'hur såg du det?' så, det är ganska mycket känslor och har lärt sig väldigt avancerat, så det är inte så att vi kan göra det heller faktiskt [förklara våra beslut]*."*

- Thomas Schön

Thomas Schön anser vidare att det är rimligt att vara orolig för att sakna insyn och förståelse för självlärda algoritmer. Samtidigt menar han att oron inte ska gå för långt då vi idag vet att deep learning, en typ av maskininlärning, fungerar och genererar de bästa resultaten inom till exempel bild- och musikigenkänning. Han liknar det med den tidiga användningen av differentialekvationer för att bland annat förutsäga planeternas rörelse, forskarna hade ingen aning om exakt hur ekvationerna löste problemen, endast att de gjorde det. Schön anser att hur dessa system fungerar är nödvändigt att förstå men oförståelsen ska inte hindra oss från att använda dem.

"Det är rimligt att vara lite orolig, det kan jag tycka, men man ska nog inte behöva vara för orolig heller utan det utvecklas ju hela tiden. Det vi vet idag är ju att deep learning levererar de absolut bästa resultaten bildigenkänning, segmentering av bilder, igenkänning av musik och generering av musik och så vidare. Vi vet ju att det funkar, det är det inget snack om. Det är snarare så att vi måste förstå ännu mer varför det funkar."

- Thomas Schön

4.2.2 Bias och representativa data

I detta tema behandlas problematiken kring bias i en AI:s inputdata och hur detta kan påverka vid uppbyggnad av en självlärd algoritm.

4.2.2.1 Juridik

Flertalet respondenter har nämnt bias som ett etiskt svårt område. Bias går inte att undvika, det finns i data som speglas av historiska förhållanden. Bias finns potentiellt också hos den som utvecklar mjukvaran.

"Det är sjukt svårt att komma från programmerarens egna uppfattningar och ska man vara ärlig så är det förstås sjukt svårt för en mänsklig domare att ställa sig utanför sina egna uppfattningar och kunna göra ett opartiskt dömande utifrån de förutsättningar och paragrafer som finns."

- Magnus Ivarsson

John Lagström lyfter också att algoritmer som bygger på dataset som innehåller bias riskerar att förstärka befintliga biaser.

"Det finns ju en massa bra exempel där man använder algoritmer för beslutsfattande som innebär att du bara förstärker det datat som algoritmen har att tillgå från början, så kallade självförstärkande algoritmer och det får man ju vara väldigt försiktig med."

- John Lagström

Andrea Lindblom understryker att bias inte är något unikt för maskinella system, utan det förekommer även vid mänsklig bedömning.

"Det som det ofta pratas om är den här rädslan att AI ska ha olika bias. Det som dock ofta missas i diskussionen är att detsamma gäller för människor i allra högsta grad."

- Andrea Lindblom

Markus Naarttijärvi lyfter även flera aspekter av självlärda algoritmer som han anser är viktiga gällande transparens.

"Problemet är ju att man bygger sina förutsägelser om framtiden på existerande data som inte nödvändigtvis är representativ för verkligheten, men framförallt är det historiskt data som inte nödvändigtvis är representativ för den framtida utvecklingen. Det tredje är att man bygger de här förutsägelserna på data från en populationsnivå eller en ganska stor grupp men sedan applicerar den på en individnivå, och den typen av kopplingar från populationsdata till en individ är väldigt svåra att göra."

- Markus Naarttijärvi

4.2.2.2 AI

Vad som också återkommer i intervjuerna med AI-respondenterna är att det finns en problematik kring representativiteten i systemets inputdata. Detta innebär att all inputdata som finns tillgänglig inte nödvändigtvis anses vara lämplig att använda, då den till exempel kan innehålla bias som inte anses vara lämplig i ett dataunderlag.

"...däremot så är den andra utmaningen mycket större, det vill säga att när världen inte är som vi vill att den ska vara, det vill säga till exempel det här med lönediskriminering, att män och kvinnor får olika lön för samma jobb och att de yrken där det är övervägande kvinnor har lägre löner än de yrken där det är övervägande män. [...] *Utmaningen där blir ju då vem som ska bestämma hur datan ska se ut, och då måste någon göra ett normativt ställningstagande och säga att "Nej, så här ska den se ut", och sen måste du tvinga världen på sätt och vis att bete sig som man önskar att den gjorde."*

- Fredrik Heintz

Samtidigt menar Fredrik Heintz att det också kan finnas ett värde i att ha data med en inbyggd bias, eftersom medvetna bias ligger till grund för vissa urval, exempelvis i antagningsprocessen till universitet.

"Vi har ju massa bias som vi vill ha. Ta det här med antagning till universitet eller till gymnasieskolan till exempel. Det bygger ju på att vi har en, diskriminering kanske är fel ord men, liksom ett urval baserat på till exempel betyg. Då har vi föreställningen att det är tillräckligt korrekt och unbiased eller vad man ska säga. Så vi gör ju massa urval hela tiden. [...] *Om det skulle vara helt unbiased då skulle det ju vara helt slumpmässigt, det vill säga att då finns det ingen struktur och absolut inget värde i datan. Det är ju strukturen du vill komma åt, det är signalen i bruset du vill komma åt."*

- Fredrik Heintz

Det har under intervjuerna även framkommit att det är viktigt att ha en stor mängd inputdata för att få en AI att ge det utfall som är önskat. Detta kan speciellt bli ett problem i mer ovanliga brottsfall, där det inte finns speciellt många tidigare fall att gå på, eller då många existerande fall plockas bort då de anses ha ett icke-önskvärt bias.

Den vanligast diskuterade etiska frågan bland respondenterna är bias. Dataset som används för att skapa och lära en algoritm kantas av mänskliga värderingar, historiska domar grundas på beslut som mänskliga domare har tagit och utvecklarens vilja kan vara av stor betydelse vid framtagning av algoritmen.

"Det kommer finnas spår av konstiga mänskliga fördomar, begränsningar, kognitiva förmågor i datat, som vi inte kan sätta fingret på när vi tittar på datat, men som kommer spela ut när en AI identifierar trender i datat. Sådana spår kommer finnas med där på något sätt."

- Jacob Dexe

"Man bör inte låta AI-forskare bestämma vad som är etiskt, utan man vill ha en domänexpert för det. Den här typen av värderingar kommer vi inte lära oss från data, utan det är krav som vi vill bygga in i modellen, och därför måste vi ta den informationen någon annanstans ifrån."

- Fredrik Johansson

En annan aspekt som diskuteras är vilka variabler som en algoritm bör undvika och vilka som bör viktas tyngre för att få en önskad output.

"Man måste bestämma sig för hur man vill att utfallet ska vara om man säger så, sen måste man balansera datat därefter. [...] *Baserar man det på fel variabler kan det bli orättvist bara därför att man har fel variabler med."*

- Josefin Rosén

"När man bygger skarpa maskinlärande modeller som till exempel neurala nätverk, i slutändan när de har fått jobba sig igenom hela nätverk så kan de definieras av miljontals parametrar, och då är det svårt att gå tillbaka och ändra och göra sig av med variabler i efterhand, för det blir sådan abstraktion allteftersom modellen byggs eller algoritmen utvecklar sig. Där är det superviktigt att man balanserat sin data, att man har koll på datat redan innan det går in i algoritmen."

- Josefin Rosén

4.2.3 Kunskap och kompetens

I detta tema förmedlas respondenternas syn på både den kunskap och kompetens som finns idag gällande AI samt vilka krav som ställs kunskaps- och kompetensmässigt för att automatisera de svenska domstolarna.

4.2.3.1 Juridik

Det har under studien även framkommit att AI som begrepp är svårtolkat och att det saknas kunskap bland många jurister. Att det saknas kunskap kan leda till att tekniken blir svårförståelig för många jurister, vilket riskerar i att resultera i att en AI inte accepteras som ett lämpligt verktyg.

"Det är få som förstår vad det [AI] *egentligen är och det är ett väldigt luddigt begrepp som väcker känslor. När jurister inte vet vad någonting är eller inte tycker att det är något för dem - då blir det väldigt lätt att man slår ifrån sig och säger att det där ska man inte hålla på med. Det tycker jag är fel inställning. Jag hade önskat att det fanns en större nyfikenhet och engagemang bland det offentliga men även bland jurister i stort att intressera sig i vad den här tekniken kan och bör göra för oss."*

- Andrea Lindblom

John Lagström lyfter risken med att kompetens tappas vid automatiseringar. Om enklare beslut automatiseras riskerar kunskapen kring dessa processer att försvinna, och om systemet skulle fallera är det inte säkert att det finns individer som besitter den kompetens som krävs för att täcka upp. Utöver denna risk så försvinner möjligheter för juniora jurister att införskaffa värdefull erfarenhet.

"Anledningen till att jag tror att man kanske ändå ska en människa på slutet som kvalitetskontrollant. Det handlar om inlärningen mest, alltså kompetensöverföringen. Det är ju idag inte de allra mest erfarna domarna som sitter och tar beslut om skilsmässor om man säger så, utan de är de nya juristerna och för dem kan det vara rätt viktigt att titta på hur en bodelning går till."

- John Lagström

4.2.3.2 AI

Även vissa av AI-respondenterna argumenterar för att en av anledningarna till att AI ännu inte har implementerats inom det svenska domstolsväsendet är att användarna inte har den kompetens som krävs för att använda verktygen till fullo, samtidigt som det finns en viss skepsis samt rädsla gentemot AI.

"Överlag i Sverige är det så att man inte har kommit så jättemycket längre än att kanske möjligtvis automatisera vissa enklare saker och liksom mer explorativt tittar på data. Det är inte så jättemycket AI som är implementerat idag, på den nivån. Man är fortfarande lite rädd och man är fortfarande lite skeptisk och har inte alltid kompetensen som kanske krävs, för att kunna använda den till fullo då."

- Josefin Rosén

4.2.4 Tillit och legitimitet

I detta tema presenteras hur allmänhetens tillit till domstolsväsendet skulle kunna påverkas vid en AI-implementering i Sveriges domstolar.

4.2.4.1 Juridik

Juridikrespondenterna nämner flera olika utmaningar som finns kopplade till tillit och legitimitet vid en AI-implementering. Bland annat tar Johan Sanner upp frågan kring förtroende för statsapparaten och dess påverkan vid en implementering.

"Utmaningar och risker med det här är naturligtvis tillit och förtroende och integritetsfrågor och såna saker. Hur mycket vill medborgarna att maskinerna ska blanda sig i? Och vilket förtroende har man för statsapparaten då? Nu har vi ju i Sverige fortfarande ganska högt förtroende för staten. [...] *Det gör kanske att det är lättare på nåt sätt, att man litar på att om nu staten har gjort det här, avvägt och balanserat, och med en folklig förankring och att det har utretts och sånt där, att nu inför man det här till exempel i domstolar så kanske man kan få mer förtroende."*

- Johan Sanner

John Lagström framför att en AI inte får ta bort det personliga mötet inom rättsprocessen. Han menar att människor idag har en positiv bild av de svenska domstolarna, mycket på grund av just det personliga mötet.

"De flesta människor som kommer i kontakt med Sveriges domstolar, de har en positiv bild av oss. De tycker att de har fått sin sak prövad och att det har gått rättvist till och sådär. Och

det tror jag handlar ganska mycket om att man tillåter det personliga mötet."

- John Lagström

Björn Östman belyser även vikten av att bli bemött korrekt.

"Ur den enskildes perspektiv så är det ju så att man vill bli sedd och man vill bli lyssnad på. Det har gjorts jättemycket bemötandeutredningar som pekar på hur viktigt det är att bli bemött på ett korrekt sätt och bli lyssnad på, så det får man inte glömma bort."

- Björn Östman

Förtroendet för domare är en annan aspekt som vore svår för en AI ta över uttrycker Magnus Ivarsson.

"Idag har ju den enskilda domaren ett högt förtroende i kraft av erfarenhet, utbildning och sin roll. Det är ett etos som är svårslaget. Det blir nog svårt, i alla fall i dagsläget, att låta en AI ta den rollen."

- Magnus Ivarsson

4.2.4.2 AI

När det kommer till att lita på AI-system så nämner Fredrik Heintz att tilliten till människor och olika professioner uppstått över årtusenden och att de nya systemen inte har funnits tillräckligt länge.

"Problematiken är att vi har ingen grund för att lita eller inte lita på de här systemen. Det vill säga, när det gäller människor vet vi hur vi själva funkar och så gör vi antagandet att troligtvis funkar de flesta som oss. Alltså om den här personen kan ge vettiga svar på de här frågorna, om den här personen har en utbildning som vi har accepterat som legitim, har klarat de här testerna och så vidare. Vi gör ju massor med saker för att man ska bli domare eller läkare eller lärare eller polis och de här yrkena. Vi har liksom över tid, genom årtusendena kunnat komma fram till någonting som gör att vi tycker att vi kan lita tillräckligt mycket på människor. Och ett, vi har inte hållit på med de här systemen tillräckligt länge för att ha haft en rimlig chans att bygga upp det där och två, att vi har en slags föreställning om

att vi människor är transparenta och att vi människor alltid fattar beslut från något slags tydligt logiskt ställningstagande, när vi återigen vet att i praktiken är det ju inte så utan vi fattar oftast beslut på lite andra grunder och sen i efterhand hittar vi på lite förklaringar, vi hittar på lite historier som verkar rimliga."

- Fredrik Heintz

Karl de Fine Licht menar att det trots allt finns fall där vi litar mer på tekniken.

"Det finns något som kallas för machine bias och det är ju att man litar på maskinen mer och det där är ju någonting som är väldigt intuitivt för de flesta av oss tror jag. Jag använder ju Google Maps, det är sällan jag tänker rent logiskt kring hur det borde funka utan jag litar på att Google Maps vet den bästa vägen."

- Karl de Fine Licht

Han tar också upp vikten av förklarbarhet för att systemen ska upplevas som legitima.

"Om man vill att folk ska uppleva att de här systemen är legitima då måste de ge ifrån sig förklaringar som folk förstår."

- Karl de Fine Licht

4.2.5 Tröghet inom det juridiska väsendet

I temat diskuteras den befintliga trögheten i det juridiska systemet, som tydliggjorts under intervjuerna, samt hur den påverkar möjligheten att implementera ny teknik.

4.2.5.1 Juridik

Flertalet respondenter har lyft att den tröghet som finns inom det juridiska systemet är en önskvärd egenskap och att förändringar bör ske långsamt.

"Juridiken ska vara lite trögrörlig, det ska inte gå för fort, för konsekvenserna om det blir fel är enorma. Ny lagstiftning ska ta tid och förändringar i sättet vi arbetar på måste ta tid, därför att det är en förutsägbarhetsfråga att folk inte gör på helt olika sätt. Då drabbas liksom hela rättsstaten av att det blir spretigt och konstigt och knepigt, och då funkar inte, då

kommer vi tappa allt vad gäller legitimitet för juridiken… Sverige är ju ett väldigt modernitetslängtande land. Vi har en strävan efter att vara moderna och vi bygger liksom alltid vårt samhälle på framtid. Vi är inte rädda för det nya."

- Anna-Sara Lind

Vidare har vissa respondenter tagit upp det juridiska systemets tröghet i förhållande till riskerna med teknisk utveckling där AI är inblandad kan gå för snabbt. De menar på att det är viktigt att en eventuell implementering av AI sker i en takt som gör det möjligt att utvärdera och analysera vilka förändringar som sker och hur de påverkar det juridiska systemet som helhet.

"...Det ställer höga krav på både oss som ska utveckla det, men även de som ska använda det, och där är jag helt på det klara med att det här är ett långlopp, det är inte någonting som ska hända över natt."

- John Lagström

"När det gäller en teknik som är väldigt hypad, där det finns väldigt mycket ekonomiska intressen, och det finns starka ekonomiska intressen som driver på för att sälja de här systemen. Och det finns potentiellt väldigt stora effekter av de här systemen så skulle jag säga att det finns ett värde, det finns ett demokratiskt och rättsligt värde i att bara sakta ner lite grann, se hur det funkar på andra ställen, inte behöva ligga i framkant, kunna se om det här är ett system som vi vill implementera, vilka effekter vi kan få och så vidare. Så på det sättet så, jag är inte skeptisk till AI som sådant, jag tror att det har väldigt stark potential inom väldigt många områden, när det gäller användandet av det i rättsväsendet … man behöver inte ligga i framkant, man kan vara försiktig och det finns uppenbara problem utifrån de principer som vi har idag."

- Markus Naarttijärvi

4.2.5.2 AI

Det har framkommit under intervjuerna med respondenterna att de problem som redan finns inom domstolarna kan förstärkas ytterligare vid en AI-implementering. Detta eftersom en implementering av AI leder till att processer går snabbare att genomföra.

"...Men att det också kommer med ett stort ansvar, för det som vi har gjort nu kan göras väldigt mycket snabbare, och blir det fel så kan det väldigt snabbt bli väldigt fel. [...] *När jag pratar på föreläsningar brukar jag citera Spindelmannen: 'With great power comes great responsibility'."*

- Josefin Rosén

"Jag tror att domstolsväsendet kommer vara en av de delar i offentlig förvaltning som kommer gå över till storskalig AI sist. Jag tror inte att vi är bekväma med de konsekvenser det kan få. [...] *Det som möjligen kommer att hända tidigare är att det kommer finnas i förundersökningar, och att man behöver ta ställning till AI-beslut i den delen av processen."*

- Jacob Dexe

4.2.6 Kritik mot befintliga system

I följande tema belyses respondenternas åsikter kring redan existerande AI-system inom offentlig förvaltning.

4.2.6.1 Juridik

Bland vissa av juridikrespondenterna diskuterades Trelleborgsmodellen. De respondenter som nämnde den var av olika skäl kritiska mot den. Anders Hagsgård menar att det finns en problematik kopplat till transparensen i Trelleborgsmodellen.

"Jag antar att ni har läst om Trelleborgsmodellen. Facket har ju klagat på det nu, att man inte vet hur algoritmen ser ut [...]*. Den borde finnas öppet tillgänglig, 'såhär ser algoritmen ut'. Sen kanske du inte förstår den, men 'kommer du hit så får du se hur koden ser ut', eller får en beskrivning av den.* [...] *Nu är det bara ett svart hål eller en svart burk, du vet inte vad som finns inuti."*

- Anders Hagsgård

Markus Naarttijärvi menar att ett annat problem med Trelleborgsmodellen är att systemet har fått beslutsbefogenheter på fel grunder.

"Det är en domare som har rätt att besluta om en dom, det är en befattningshavare hos en myndighet som har rätt att fatta beslut och så vidare. Den kan man väl säga förutsätter en grad av självbestämmande hos den personen. Den är svår att översätta till en AI-kontext. Jag vet att till exempel Trelleborgs kommun när de började använda AI-verktyg har hävdat att de gör de på basis av att det här en digital arbetstagare, och att man då kan delegera beslutsbefogenheter till den precis som en människa, för att den är bara en digital arbetstagare så att säga, och det är på alla sätt fel."

- Markus Naarttijärvi

4.2.6.2 AI

Även bland några av AI-respondenterna har det resonerats kring de redan existerande systemen inom offentlig verksamhet. Frank Dignum argumenterade för varför han ansåg att Trelleborgsmodellen inte har varit speciellt framgångsrik.

"A system like that might ask you 'How many children do you have living at home?' Is it my biological children? Is it the children that actually live with me all the time or children that every other week are with me, because I am divorced, and my children live half the time with the other partner? So, depending on the answer and how you interpret that kind of questioning and give information, the system might decide on giving welfare or not, without you being aware that maybe there are mistakes there. And this is the kind of things that I know is quite subtle."

- Frank Dignum

Även det i USA frekvent använda systemet för att prediktera återfallsrisk, COMPAS, har vid något tillfälle nämnts. Det har till exempel diskuterats om hur ett resultat från COMPAS påverkar domaren som dömer i det fallet där den har använts.

"När man använder COMPAS till exempel har man en tendens att förstärka den första magkänslan som man hade kring en person. Om du är negativt inställd till någon som kommer in, och så säger vi att den (COMPAS) ger en sjua av tio, då tenderar man att bli ännu mer säker på sin sak. Om man har en magkänsla som säger att 'det där är ingen farlig person', och så får den sju av tio, då tenderar du i ännu högre utsträckning att lita på din

magkänsla. [...] Då måste man designa och förpacka informationen på ett sätt så att man undviker den typen av bias."

- Karl de Fine Licht

4.3 Implementeringsaspekter

I följande tema presenteras identifierade faktorer att ha i åtanke vid implementering av AI i svenska domstolar, samt mer konkret inom vilka områden det kan användas och i sådant fall hur.

Tabell 6: I tabellen presenteras exempelcitat från området Implementeringsaspekter.

Tema	**Exempelcitat**
Användningsområden	*"Jag tror att man kan använda AI till vissa former av sortering, att hitta data, lyfta fram data, den typen av arbete. Egentligen fram till någon form av beslutsprocess där ser jag begränsningarna idag."* - Magnus Ivarsson
Placering och fördelning av ansvarsbördan	*"Det är en obekväm sits för en mänsklig beslutsfattare som är slutligen ansvarig för ett beslut om den inte själv känner att den har förmåga att förstå, ifrågasätta, komma till en motsats/slutsats utan istället bara blir som en galjonsfigur som sitter och stämplar de här besluten och ändå blir ansvarig för utfallet."* - Markus Naarttijärvi
Prestandakrav på AI	*"Vi kräver ju inte perfektion ens av människan så det skulle vara väldigt svårt att kräva det av en AI. Ska vi genomföra en förändring ska det var en förändring som leder till en förbättring så varför ska man inte ställa högre krav på systemet än på människan. Ska*

	vi ha ett system får det väl prestera bättre än människan." - Magnus Ivarsson
Möjlighet till överklagan	*"Så länge du kan överklaga beslutet så är det inte så farligt, då kan man automatisera rätt många såna här beslutsprocesser. Jag menar vissa saker kan vi inte ens överklaga idag utan en åklagare kan ta beslut gällande vissa saker."* - Karl de Fine Licht
Metodik för implementering och lämpliga metoder	*...det är en typisk bildbehandlingsuppgift, då ska vi självklart använda oftast neuralnätsbaserade metoder som är det som funkar bäst där. Är det däremot så att vi har de här regelverken och nu ska vi automatisera regelverken, troligtvis så är det bättre att ha någon mer, ja, logikbaserad eller regelbaserad metod för det… Det beror helt på vad det är för problem man vill lösa, sen får man hitta eller välja teknik därefter."* - Fredrik Heintz

4.3.1 Användningsområden

I detta tema presenteras inom vilka områden av domstolsväsendet som respondenterna finner det lämpligt att implementera AI.

4.3.1.1 Juridik

Enligt juridikrespondenterna finns det möjlighet att implementera AI som beslutsstöd för domstolarna. En AI skulle i ett sådant fall exempelvis kunna samla in och sortera relevant data för det aktuella målet.

"Jag tror att man kan använda AI till vissa former av sortering, att hitta data, lyfta fram data, den typen av arbete. Egentligen fram till någon form av beslutsprocess där ser jag begränsningarna idag."

- Magnus Ivarsson

Några respondenter menar att vissa typer av beslut, exempelvis skilsmässor, kan fattas av en AI. Det ska ändå påpekas att dessa respondenter trycker på att detta endast får ske i de fall där det finns en överenskommelse att använda AI eller fall som är helt binära.

"Det finns inte så många måltyper som är helt binära. De flesta äktenskapsskillnader är dock binära. Har parterna barn under sexton år – en uppgift som man lätt kan hämta från folkbokföringen – ska det vara betänketid och när man har haft betänketid finns det en rätt till skilsmässa. När de objektiva förutsättningarna väl är uppfyllda har den part som vill skiljas således rätt till det oavsett vad den andra parten tycker."

- Anders Hagsgård

Samtliga respondenter ställer sig däremot tveksamma till AI:s möjligheter att ersätta det mänskliga dömandet, speciellt i komplexa fall.

"Att man i närtid skulle ersätta det mänskliga dömandet med AI, det känner jag mig tveksam till."

- Anders Hagsgård

John Lagström lyfter de problem som AI får angående att sätta in data i ett sammanhang, något som människor idag är mycket bättre på.

"Nackdelarna här är ju det att som är svårt att uppnå med AI och det är det här med sammanhang. Att kunna förstå ett sammanhang och sätta in omständigheter i det. Det är därför jag inte tror på att ersätta det mänskliga dömandet med en dator. Däremot tror jag att det finns en stor potential att höja kvaliteten på det mänskliga dömandet med hjälp av en dator."

- John Lagström

Vad som också nämnts under vissa intervjuer är att en AI kan ha svårigheter i att bedöma trovärdighet i muntliga framställningar. En människa kan tyda kroppsspråk och andra typer av signaler, vilka också är viktiga vid ett framställande. Dessutom nämns det att det är enklare för en människa att värdera bevisfrågor jämfört med en AI.

"Då är det en bevisfråga 'Har du bevisat det du påstår eller inte?'. [...] *Det har jag väldigt svårt att se, att man skulle förlita sig på en AI som säger 'Vi tror inte på den personen'..."*

- Anders Hagsgård

4.3.1.2 AI

Även bland respondenterna med AI-expertis anses beslutsstöd vara ett av de primära användningsområdena. Vissa av AI-experterna anser även att en AI kan ta egna beslut i några specifika fall.

"...kunna ta fram underlag för en människa i domstolen till exempel, att kunna ta bättre beslut och att kunna ta sina beslut mer effektivt. [...] *Kan vara allting från att prediktera, förutsäga olika saker baserat på data* [...]*."*

- Josefin Rosén

Thomas Schön nämner att analys av text är en vanligt förekommande arbetsuppgift inom domstolsväsendet. Han menar att detta är en arbetsuppgift där en maskin kan ha en bättre förmåga att utföra än vad en människa har, i vissa avseenden. Detta gäller även för analys av bild och video, vilket gör det oetiskt att inte ta hjälp av AI för dessa arbetsuppgifter.

"Det är väldigt mycket texter som processas på olika sätt, man letar efter komplicerade mönster på olika vis. Att där då ha stöd av AI på olika sätt känns ju som en fantastiskt bra idé tycker jag, för att vi människor klarar inte av att läsa så himla mycket. [...] *Både då analys av ljud och bild och text, där då maskinerna faktiskt, när det kommer till att hitta saker, så att säga, är duktigare än vad vi är. Så att det är ju ganska oetiskt att inte använda sig av det då, utan att använda begränsade mänskliga hjärnor för en sådan avancerad uppgift."*

- Thomas Schön

4.3.2 Placering och fördelning av ansvarsbördan

I detta tema diskuteras kring vem som har ansvaret för de resultat som en AI åstadkommer samt för de beslut som tas av en AI.

4.3.2.1 Juridik

Samtliga respondenter anser att det måste finnas ett mänskligt ansvar för de beslut eller resultat som genereras av en AI. Detta faller huvudsakligen på användaren. Mellan olika respondenter råder det inte konsensus om vem som är användaren, vissa hävdar att ansvaret ligger på rätten, andra att det ligger på en specifik domare.

Markus Naarttijärvi diskuterar problematiken med att en domare har ansvar över en AI som den inte kan påverka, vilket försätter denne i en obekväm sits.

> *"Det är en obekväm sits för en mänsklig beslutsfattare som är slutligen ansvarig för ett beslut om den inte själv känner att den har förmåga att förstå, ifrågasätta, komma till en motsats/slutsats utan istället bara blir som en galjonsfigur som sitter och stämplar de här besluten och ändå blir ansvarig för utfallet."*
>
> - Markus Naarttijärvi

Stanley Greenstein[1] spekulerar emellertid att det eventuellt skulle kunna vara möjligt att skapa en juridisk persona för en AI som då skulle kunna ta någon form av juridiskt ansvar. Han påpekar att detta endast är spekulativt och att det finns stora risker och farligheter med att lägga ansvar hos en AI. Vidare menar Greenstein på att människor inte ska kunna slippa undan juridiskt ansvar genom att gömma sig bakom en AI och att det i slutändan ändå måste finnas ett mänskligt ansvar.

4.3.2.2 AI

Det råder samstämmighet bland de tillfrågade respondenterna kring att placering och fördelning av ansvar är ett svårt problem att lösa. Respondenterna är också överens om att det

[1] På grund av tekniska problem var det inte möjligt att spela in intervjun med Stanley Greenstein. Därför är han inte direktciterad, då endast skriftliga anteckningar från intervjun finns tillgängliga.

inte bör läggas något ansvar på själva AI:n. Istället ska ansvaret fördelas på användaren och utvecklaren enligt framtagna avtal och policys.

"Såsom saker är uppbyggda är det bara vi människor som kan ta ansvar."

- Fredrik Heintz

"Går man sen så att man tar bort hela människan där, då blir det mycket mer komplicerat hur man ska göra. Är det då den som har byggt algoritmen eller är det företaget eller kan man ha någon ny form av juridisk enhet som det gäller vissa regler för?"

- Thomas Schön

Dessutom hävdar flera av respondenterna att det är viktigt att ansvaret utkrävs på ett effektivt och tydligt sätt, så att den ansvarige för AI:n inte har möjlighet att skylla ifrån sig.

"*I många fall så är det svårt att peka på en individ, och där tycker jag att vi har en utmaning att, vårt ansvarsutkrävande, för att det ska vara effektivt om vi säger så, måste nästan falla tillbaka på en enskild individ. Om man tar en domstol till exempel ... det är tre personer i rummet, du vet att någon av dem, minst en utav dem har slagit ihjäl personen men du kan inte peka ut exakt vem det var, och då måste du frikänna alla."*

- Fredrik Heintz

4.3.3 Prestandakrav på AI

I detta tema lyfts respondenternas diskussion och åsikter kring vilka prestandakrav som bör finnas på en AI inom Sveriges domstolar.

4.3.3.1 Juridik

Något som diskuterats av flera respondenter är att människor inte är helt perfekta och att då sätta perfektion som standard för att införa AI är inte lämpligt. Den slutsats som samtliga respondenter som diskuterat denna aspekt har dragit är att kravet bör vara att en AI bidrar till en förbättring. Exakt hur mycket bättre en algoritm skulle behöva vara saknas konkreta svar på.

"Vi kräver ju inte perfektion ens av människan så det skulle vara väldigt svårt att kräva det av en AI. Ska vi genomföra en förändring ska det var en förändring som leder till en förbättring så varför ska man inte ställa högre krav på systemet än på människan. Ska vi ha ett system får det väl prestera bättre än människan."

- Magnus Ivarsson

"Nej det måste det inte [svar på om systemet måste vara perfekt] *för vi människor är inte 100-procentiga. Däremot så vi måste vi säga att om konfidensen är tillräckligt låg så ska vi inte tillmäta det någon betydelse men om den är tillräckligt hög då får domaren tillmäta det betydelse.*"

- John Lagström

4.3.3.2 AI

Även några av respondenterna från AI-sidan har diskuterat vilka krav som ska ställas på prestandan hos en AI. En AI behöver inte vara perfekt för att implementeras, det räcker med att den överträffar den mänskliga standarden och därmed leder till en förbättring. Exakt hur mycket bättre anses vara en svår fråga som i dagsläget inte går att svara på direkt. Thomas Schön diskuterar även kring att en AI förmodligen behöver vara avsevärt mycket bättre än människan för att vi ska vara bekväma med att använda den.

"Om vi pratar om till exempel en bil, hur mycket bättre måste en autonom bil vara än en genomsnittlig mänsklig förare för att vi ska tycka att den är okej? Jag tror att den måste vara avsevärt mycket bättre. Jag tror inte det räcker med att den är 20 procent bättre utan den ska vara bra mycket bättre."

- Thomas Schön

Josefin Rosén lyfter också att ny teknik och den ökade beräkningskraft som den innebär riskerar att förstärka de fel som vi för över till algoritmen. Detta är ytterligare en anledning till att ställa höga krav på den teknik som används.

"...Men att det också kommer med ett stort ansvar, för det som vi har gjort nu kan göras väldigt mycket snabbare, och blir det fel så kan det väldigt snabbt bli väldigt fel. [...] *Vi*

tenderar ju att ställa högre krav på ett AI-system än vad vi gör på en människa. Det ligger ju i det att man inte har samma tilltro i ett system som man har till en människa. Med all rätt så krävs det ju faktiskt att vi har högre precision hos ett AI-system[...] *Den behöver göra rätt liksom, för att det ska vara och kännas okej."*

- Josefin Rosén

4.3.4 Möjlighet till överklagan

I temat diskuteras hur ett beslut fattat av en maskin skulle kunna överklagas till en mänsklig instans, och hur detta skulle kunna möjliggöra användandet av AI som mer än enbart beslutsstöd.

4.3.4.1 Juridik

Johan Sanner diskuterar kring betydelsen av att kunna få ett beslut som fattats maskinellt överprövat av en människa, och lyfter möjligheten att i högre utsträckning nyttja AI i en första instans, i syfte att effektivisera instansen. Han understryker emellertid att beslut fattade i en sådan instans måste kunna överklagas.

"Det är viktigt samtidigt att det går att få det överprövat på ett bra sätt, fullföljden där alltså, om det blir en massa tokiga beslut för att en maskin sitter och spottar ur dem så måste man ju kunna få det överprövat. [...] *Man hade ju möjligen kunnat tänkt sig någon form av första instans som hade en betydande grad av maskinell inblandning, men när det lyfts upp, till exempel till Kammarrätten, så kanske man där säger att 'Här ska det iallafall vara minst 75 procent mänsklig inblandning'."*

- Johan Sanner

Stanley Greenstein[2] understryker att det vid implementering av AI är viktigt att det finns utrymme för individer som hamnat på fel sida av algoritmen att överklaga beslut. Om 98 av 100 fall bedöms korrekt måste det finnas utrymme för resterande två att kunna få sin sak prövad. Detta skulle kunna göras av en människa, men det finns inga garantier för att en människa kommer att förstå det ursprungliga beslutet som AI:n fattade. I och med detta är

[2] På grund av tekniska problem var det inte möjligt att spela in intervjun med Stanley Greenstein. Därför är han inte direktciterad, då endast skriftliga anteckningar från intervjun finns tillgängliga.

risken att det blir ett system som övervakar ett annat system, som säkerligen också har sina brister, och att den mänskliga insynen minskar ännu mer.

4.3.4.2 AI

Även några av respondenterna från AI-sidan har lyft vikten av att det går att överklaga till en människa. Frank Dignum lägger mycket vikt på att det ska gå att bestrida en dom eller ett beslut påverkat eller taget av en AI.

> *"I think it's actually good to use it* [AI] *in all those kinds of clear cut cases because it frees up the time to spend on more important cases* [...].*Given that people always can contest the outcome* [...] *So that makes it like it's not an automatic thing that always takes over."*
>
> - Frank Dignum

Karl de Fine Licht lyfter att möjligheten att kunna överklaga kommer påverkas av individens socioekonomiska bakgrund och att de personer som redan idag är utsatta kan få problem med ett mer automatiserat rättsväsende.

> *"Så länge du kan överklaga beslutet så är det inte så farligt, då kan man automatisera rätt många sådana här beslutsprocesser. Jag menar vissa saker kan vi inte ens överklaga idag utan en åklagare kan ta beslut gällande vissa saker. Ett problem med det då, till exempel om man tar socialbidrag, det är ju rätt utsatta grupper söker det såklart. De utsatta grupperna, säg att de får ett automatgenererat beslut. Det står att man kan överklaga men det är väldigt liten sannolikhet att de faktiskt kommer att göra det, det är ju så det ser ut. De har inte kapitalet nog för att göra detta. I praktiken då så kommer det bli så att de som har det tuffast är de som kommer drabbas hårdast i sådana system."*
>
> - Karl de Fine Licht

4.3.5 Metodik för implementering och lämpliga metoder

I detta tema diskuteras hur en AI-implementering inom domstolsväsendet skulle kunna gå till, och vilka metoder som lämpar sig bäst för användning.

4.3.5.1 Juridik

Bland de juridikrespondenter som har berört frågan kring hur en implementering bör genomföras, finns det en samstämmighet kring att det är viktigt att testa AI i verksamheten för att se hur det fungerar.

"AI är inte en teoretisk uppgift eller någonting som man löser bakom ett skrivbord, utan man måste komma igång och testa för att se dels vad som går att göra, vad man ska göra och om det är lämpligt att göra det på det sättet."

- John Lagström

Andrea Lindblom håller med John Lagström i att det är viktigt att komma igång med att testa hur en AI lämpar sig i domstolarna. Hon trycker även på att det är viktigt att börja med experimenten i en liten skala.

"Vi behöver komma igång på en annan nivå, på en strategisk nivå, och var börjar vi och så? Sen att man liksom låter de domstolar som är intresserade och ligger lite mer i framkant, att man kanske har småskaliga projekt som syftar till att kunna växa om det blir framgångsrikt, men att man helt enkelt testar lösningar. [...] *Jag tror inte alls på stora, bombastiska projekt som bara kostar fruktansvärt mycket pengar, utan jag tror mer på småskalighet, god omvärldsbevakning och att man vågar testa."*

- Andrea Lindblom

4.3.5.2 AI

Under intervjuerna har AI-respondenterna både diskuterat kring viktiga element att ha i åtanke vid implementeringen, och nämnt flera konkreta metoder som skulle kunna implementeras för att assistera med frekvent förekommande arbetsuppgifter. Josefin Rosén understryker att en så precis algoritm som möjligt bör användas, men att vikten av förklarbarhet inte får förringas. Rosén betonar också att det måste ske en kontinuerlig granskning av applikationer för att garantera att kvaliteten bibehålls.

"Jag tycker att man ska använda så bra algoritmer, med så hög precision som möjligt, så länge man har möjlighet att förklara dem. Funkar det lika bra med en väldigt enkel modell så ska man såklart använda den, annars tycker jag att man kan bara se till att man har

förklarbarhet och transparens… och att man inte bara släpper applikationen fri att göra vad den vill utan även om man har automatiserat saker så måste man hela tiden monitorera och även göra audit eller revision regelbundet på det då."

- Josefin Rosén

Thomas Schön nämner att maskininlärning för textanalys är ett intressant område att titta på och att utvecklingen går snabbt där, men att det kanske fortfarande inte är redo för användning.

"[Om att använda Maskininlärning för textanalys] *Det är väl en bit kvar, det händer mycket där, så kan jag säga. Det har hänt jättemycket de senaste fem åren bara, både när det gäller att analysera befintlig text, men också när det gäller att skapa text, alltså skriva text, har det hänt väldigt spännande grejer senaste 2 åren faktiskt. Så det pågår en väldigt kraftig teknikutveckling där, sen om någonting av det som finns klart är redo att användas nu för det, det vågar jag faktiskt inte säga för jag är för dåligt insatt i detaljerna. Men det känns som en väldigt bra sak att titta på om man skulle vara intresserad av det där, det känns inte omöjligt."*

- Thomas Schön

Fredrik Heintz lyfter att det inte bör finnas en bestämd metod för implementeringen utan att den bör anpassas efter det problem som ska lösas. I intervjun nämner han neurala nätverk och mer logikbaserade metoder som lösningsexempel på olika problem som kan uppstå inom det juridiska väsendet.

"Vi försöker välja metod efter problem, snarare än att använda samma metod på allting och så löser vi bara de problem vi kan lösa med det. Därför så beror det ju på, om det nu är till exempel, vi hade någon session tillsammans med polisen, och det var väl till exempel det här med att leta föremål, om du nu hade hittat typ någon jacka eller någonting på någon plats som du kunde knyta till brottslingen, kan vi nu hitta samma jacka i tusentals timmar film eller tusentals foton, det är en typisk bildbehandlingsuppgift, då ska vi självklart använda neuralnätsbaserade metoder som är det som funkar bäst där. Är det däremot så att vi har de här regelverken och nu ska vi automatisera regelverken, troligtvis så är det bättre att ha

någon mer, ja, logikbaserad eller regelbaserad metod för det… Det beror helt på vad det är för problem man vill lösa, sen får man hitta eller välja teknik därefter."

- Fredrik Heintz

4.4 Respondenternas inställning

I detta tema redogörs respondenternas generella inställning till AI i svenska domstolar.

Tabell 7: I tabellen presenteras exempelcitatet från området Respondenternas inställning.

Tema	**Exempelcitat**
Respondenternas inställning	*"Jag tycker att det ska vi göra* [implementera AI] *vi har en skyldighet att handlägga målen så kostnadseffektivt som möjligt och fortsätta utveckla vår verksamhet. Ser vi att det kommer verktyg som skulle kunna hjälpa oss, om än idag i begränsad mängd, så är det vår förbannade plikt och skyldighet att undersöka på vilket sätt vi kan använda oss av AI."* - Magnus Ivarsson
	"Borde vi inte ha rätten att bli dömd av en maskin i stället, med avseende på det här med att vi faktiskt kan använda den för att minska subjektiviteten, öka objektiviteten och öka sannolikheten att det faktiskt är så korrekt det kan bli." - Fredrik Heintz

4.4.1 Juridik

En majoritet av respondenterna är åtminstone i någon mån positivt inställda till en implementering av AI i Sveriges domstolar, om än en försiktigt sådan. Det är ingen av respondenterna som ställer sig helt emot AI inom domstolarna, dock varierar åsikterna om hur långt en implementering bör gå och inom vilka områden en AI kan tillämpas.

"Jag tycker att det ska vi göra [implementera AI]*, vi har en skyldighet att handlägga målen så kostnadseffektivt som möjligt och fortsätta utveckla vår verksamhet. Ser vi att det kommer*

verktyg som skulle kunna hjälpa oss, om än idag i begränsad mängd, så är det vår förbannade plikt och skyldighet att undersöka på vilket sätt vi kan använda oss av AI."

- Magnus Ivarsson

Anders Hagsgård nämner också att samhället förändras och att vi kommer behöva lösa uppgifter på andra sätt än tidigare.

"Det handlar i slutändan om skattebetalarnas pengar, vi får en äldre och äldre befolkning och färre som är yrkesverksamma så vi måste lösa uppgifter på ett annat sätt. Så rätt använt är jag positiv till det."

- Anders Hagsgård

Johan Sanner framför också betydelsen av att en sådan implementering har ett faktiskt syfte när den väl sker.

"Vi vill ju att det ska bli ett bättre samhälle, vi ska hushålla med resurser, bättre kvalité, kvantitet, besked i rätt tid och allt det här. Om man ska börja med något sådant här så måste man också kunna påvisa de samhälleliga vinsterna och kopplingen till demokrati och rättsstat, de här tunga sakerna som är viktigt. Annars blir det en ingenjörsprodukt som inte har någon förankring, utan man måste få med sig hela det här om man ska komma vidare tror jag."

- Johan Sanner

4.4.2 AI

Bland AI-respondenterna fanns en varierad inställning till en implementering av AI inom de svenska domstolarna. Överlag var dock majoriteten av de tillfrågade positivt inställda.

"*Är det någonting som är väldigt svart eller vitt så skulle man helt kunna automatisera det, men allra oftast så blir det ju bäst när man låter maskinen och människan arbeta tillsammans, och då handlar det mycket om att kunna ta fram underlag för en människa inom domstolen."*

- Josefin Rosén

"Helt emot … Jag vet visserligen inte i vilken utsträckning det svenska domstolsväsendet begår systematiska fel idag, men jag känner inte till någon aspekt där det måste vara AI som löser effektiviseringsfrågan, istället för att förändra arbete eller process på andra sätt. AI är inte moget nog för att alla problem absolut måste lösas med just den tekniken snarare än med andra alternativ."

- Jacob Dexe

"Borde vi inte ha rätten att bli dömd av en maskin i stället, med avseende på det här med att vi faktiskt kan använda den för att minska subjektiviteten, öka objektiviteten och öka sannolikheten att det faktiskt är så korrekt det kan bli."

- Fredrik Heintz

5. Diskussion

I detta kapitel diskuteras rapportens resultat och analys utifrån det teoretiska ramverket. Diskussionen sker utifrån fem huvudområden; *Direkta effekter av AI-implementering, Indirekta effekter av AI-implementering, Tekniska aspekter, Tillvägagångssätt vid implementering* och *Vidare forskning och utredning*.

5.1 Direkta effekter av AI-implementering

Det är utifrån intervjuerna möjligt att dra slutsatsen att en implementering av AI på flera sätt skulle vara positivt för det svenska domstolsväsendet. En av de mest betydande fördelarna är att domstolsväsendet skulle kunna bli mer resurseffektivt, vilket kan resultera i snabbare och mindre resurskrävande rättsprocesser. Detta skulle kunna generera stora finansiella besparingar inom domstolarna, vilket gör att ekonomisk tillväxt gynnas eftersom resurser kan omfördelas och nyttjas mer effektivt. På så vis främjas ett av FN:s globala mål för hållbar utveckling, det som berör anständiga arbetsvillkor och ekonomisk tillväxt.

I dagsläget går mycket av resurserna åt till att göra enkla, repetitiva uppgifter som en AI istället skulle kunna utföra. En sådan uppgift som förekom frekvent under intervjuerna med juridikrespondenterna var behandlingen av skilsmässor. Dessa processer är oftast helt binära, vilket innebär att det kan vara lämpligt för en AI att hantera sådana typer av mål. Samtidigt ska det beaktas att dessa typer av mål ofta är en del av skolningen av juniora jurister. Därmed finns det en långsiktig risk att en AI-implementering kan verka negativt på deras kompetensnivå.

Implementering av AI inom domstolsväsendet har också potential att jämna ut de ojämlikheter som finns i domstolsväsendet idag. En av grundpelarna i ett rättssystem är att inga skillnader görs mellan olika individer med avseende på bland annat kön, etnicitet och ställning i samhället. Det har tidigare i rapporten emellertid nämnts att inkonsekventa bedömningar förekommer i Sverige. En korrekt implementerad AI skulle kunna vara behjälplig för att jämna ut dessa skillnader, och skulle förhoppningsvis kunna leda till att förbättra möjligheterna att uppnå hållbarhetsmålen kring minskad ojämlikhet och jämställdhet i det juridiska systemet. Samtidigt finns utmaningar relaterade till dessa mål, en

AI som är tränad på tidigare domar kan påverkas av om de personer som dömde i dessa mål hade någon typ av oönskad bias. I de fallen behöver någon besluta om vilka gamla domar med bias som systemet ska tränas på, och vilka som ska förkastas. Eftersom detta är ett mänskligt beslut att ta finns det även här en risk att bias smyger sig in, då den person som tar beslutet eventuellt väljer ut de domar som denne personligen anser är korrekta.

För att det ska vara aktuellt att implementera en AI i domstolsväsendet måste det leda till att en kvalitetsförbättring sker. Hur mycket bättre den behöver vara är emellertid en svår fråga att svara på, och är något som kan variera beroende på vilken uppgift det är som en AI ska utföra. I en roll där en AI är med och fattar ett beslut eller agerar beslutsstöd är det inte säkert att en liten förbättring räcker för att en AI ska accepteras. Det finns även en risk att allmänhetens bild av domstolsväsendet försämras om den mänskliga kontakten försvinner. Detta eftersom en AI-implementering kan leda till att de individerna som kommer i kontakt med en domstol kan känna att de inte blir sedda och hörda på samma sätt som tidigare. En implementering av AI som leder till en kvalitetsförbättring inom de svenska domstolarna bidrar till att FN:s delmål kring att främja rättssäkerhet och säkerställa tillgång till rättvisa uppnås, eftersom ett bättre domstolsväsende induktivt korrelerar med en ökad rättssäkerhet.

5.2 Indirekta effekter av AI-implementering

Under intervjuerna har det framkommit att transparens- och förklarbarhetsproblematiken är mycket viktig att ta itu med, vilket även stärks av de befintliga etiska riktlinjerna för implementering av AI inom domstolsväsendet. Ur ett allmänt mottagarperspektiv blir det sannolikt inte tydligare om all kod finns tillgänglig. Däremot är tillgång till information som ger en tydlig förklaring av logiken bakom beslutet ytterst nödvändig. Detta i syfte att skapa och säkerställa tillit till systemet. Rimligtvis bör förklaringarna motsvara eller överträffa de som ges i dagens domstolar utan AI-inblandning. Samtidigt bör det understrykas att människor inte är särskilt transparenta och inte sällan har problem att motivera sina beslut, vilka ofta fattas baserat på känsla och erfarenhet. Tilliten till det mänskliga dömandet härstammar snarare från tilltron till statsapparaten, och vetskapen om att individerna som fattar besluten också har en gedigen utbildning, än från förklaringar som människor ger.

Om AI-verktyg skulle implementeras i svenska domstolar uppstår en annan problematik som lyfts av studiens respondenter med juridisk expertis, nämligen att källkoden enligt

offentlighetsprincipen är en allmän handling. Att göra systemet fullständigt transparent kan medföra risker. I fel händer kan källkoden vara mycket destruktiv, med rätt kompetens finns möjlighet att i någon mån exploatera systemet. Lägg dessutom till att i fall där algoritmen ska tränas på ett dataset baserat på exempelvis historiska domar, så måste datasetet också vara offentligt för att uppnå fullständig transparens. Bland informationen som återfinns i dataseten finns troligen uppgifter som är av känslig natur och normalt skulle sekretessbeläggas. Sekretessen skulle förmodligen kunna bevaras, samtidigt som det är värt att nämna att när en algoritm utvecklas och tränas så görs det antagligen i samarbete med en extern mjukvaruutvecklare, vilket kräver att känsliga data delas. Detta ställer mycket höga krav på dataskydd hos båda parter, eftersom det i fall där dessa känsliga uppgifter läcker ut till allmänheten kan ha allvarliga konsekvenser.

En annan aspekt i transparensfrågan är att ett verktyg måste kunna kontrolleras och testas för att garantera dess kort- och långsiktiga kvalitet. Så länge verktyget fungerar på en nivå som är jämförbar med eller bättre än ett mer traditionellt tillvägagångssätt så är detta inte särskilt problematiskt, men skulle verktyget inte fungera kan problemet bli mycket svårlöst om det exempelvis döljer sig i ett dolt lager i algoritmen. För att kunna introducera ett AI-verktyg på ett säkert sätt behövs därför experter som med hög precision kan validera koden och förstå logiken bakom denna.

Den mest frekvent diskuterade etiska frågan vad gäller AI-implementering inom domstolsväsendet berör bias. Detta är någonting som existerar även i det befintliga systemet, eftersom människor generellt sett är partiska. Förekomst av bias i ett AI-verktyg som implementeras i svenska domstolar är oundviklig. Likaså kommer den data som utgör underlaget som en algoritm tränas på att vara färgad av historiska beslut och mänskliga värderingar. En av de största utmaningarna vid en sådan implementering är att det historiska underlaget måste vara representativt och medföra att algoritmen är icke-diskriminerande. Men vem bestämmer egentligen vad som är representativt, vem ska tolka lagen och de historiska domarna som ska skapa underlaget? Förslagsvis lämpar sig en domänexpert bättre för uppgiften än en mjukvaruutvecklare eller en AI-forskare, då denne tenderar att ha bättre koll på de externa krav som måste implementeras i en AI för att den ska gå att använda inom domstolsväsendet.

Vidare finns ytterligare problematik i att världen inte alltid ser ut som vi vill att den ska göra. Frågan är om underlaget ska baseras på den värld vi faktiskt lever i, med samhällets befintliga normer och värderingar, eller om det ska utformas med utgångspunkt i att världen är perfekt. Vem bestämmer då vad som menas med perfekt, och hur kan dataunderlag skapas som reflekterar en värld som vi inte lever i?

Att någonting har bias behöver inte heller automatiskt betyda att det är dåligt. Som belyses i resultatet finns någon form av bias i nästan alla urval som görs, i annat fall blir urvalet helt slumpmässigt och därmed i många fall intetsägande.

När det kommer till ansvar för en AI och dess output så har intervjuerna givit att det är viktigt med ett mänskligt ansvar och att AI:n i sig inte kan vara ansvarig. Inom området har det lyfts att användaren har ett stort ansvar, men även att ett visst ansvar ligger på utvecklaren. Respondenter på AI-sidan lyfte att en fördelning av ansvar skulle kunna ske med hjälp av policys och avtal. Människan ska inte heller kunna gömma sig bakom en AI för att friskriva sig från ansvar. Det finns mycket forskning att göra kring hur ansvarsfrågan bör hanteras i en värld där AI implementeras inom fler och fler områden. Är det till exempel rätt att kräva en domare på ansvar för en algoritms utfall, om denne inte själv har förmåga att förstå eller ifrågasätta utfallet? I och med den påverkan ett beslut i rättsväsendet kan ha på individen, exempelvis i form av ett utdömt fängelsestraff, så är det av yttersta vikt att vara tydlig med vem som bär ansvar. Ett effektivt ansvarsutkrävande förutsätter nämligen att det finns någon som kan ställas till svars för eventuella felaktigheter i processen.

Möjligheten att överklaga ett beslut fattat av en algoritm har lyfts av både AI- och juridikrespondenter. Karl de Fine Licht nämner att så länge det finns möjlighet till överklagan, så finns det utrymme att automatisera beslutsprocesser i relativt stor utsträckning. Han lyfter också att utsatta grupper i samhället drabbas av ett system där det är svårt och kostsamt att överklaga ett beslut. Därmed skulle denna process behöva vara enkel och lättillgänglig, för att inte riskera att dessa grupper mister möjligheten att överklaga och således blir alltmer utsatta. Samtidigt finns det en risk att alla som får ett beslut de inte är nöjda med väljer att överklaga, då det inte krävs någon insats från deras sida. Hantering av en excessiv mängd överklaganden riskerar att vara mycket tidskrävande och medför att den tilltänkta effektiviseringen av processen uteblir.

Vad som också är värt att nämna när en AI-implementering i svenska domstolar diskuteras är den tröghet som är vanligt förekommande i domstolsväsendet idag. Domstolsväsendet har ofta en konservativ inställning till att introducera ny teknik, eftersom det kan få allt för stora negativa konsekvenser om någonting skulle gå fel. Samtidigt är AI-utveckling ett område som går mycket snabbt framåt, vilket kan innebära att resultaten av att implementera AI i liknande verksamheter inte alltid har analyserats fullt ut. Det är därför av betydande vikt att en AI-implementering i domstolsväsendet inte stressas fram, utan att det analyseras och testas i en takt och skala där lagstiftning och samhället i stort kan hänga med. En aspekt som är intressant att betrakta är hur en samhällskris påverkar domstolsverkets tröghet. En kris likt den coronaviruset medfört skulle eventuellt göra att förändringar som säkerställer domstolens fortsatta förmåga att hantera och slutföra mål genomförs betydligt snabbare.

5.3 Tekniska aspekter

I rapporten har två olika typer av AI introducerats, upplärd och självlärd. En upplärd AI bör vara enklare och mindre kontroversiell att implementera i det svenska domstolsväsendet eftersom den till viss del undviker problematiken kring bias och transparens, som ett resultat av att koden är helt genererad av människor. Samtidigt så har utvecklaren av koden en egen bias som fortfarande i någon mån kommer påverka hur AI:n kommer utformas. Även transparensen kring ett sådant system bör vara bättre då skaparen själv har bestämt hur modellen ska hantera data. Exempel på sådana system är Richard Susskinds (2019) AI för preskriptionstid eller Trelleborgsmodellen.

En implementering av upplärda algoritmer kräver däremot fortfarande försiktighet och noggrannhet. Trelleborgsmodellen är på inga sätt en perfekt lösning, utan den har utsatts för kritik. Några respondenter kritiserar bland annat modellen för att vara en svart låda och att ett sådant ansvar inte kan överlåtas på ett verktyg. Alltså är det oavsett typ av AI av största vikt att tänka på hur en AI implementeras och vilka effekter det får. Det är också viktigt att belysa att självlärd och upplärd AI är olika typer av verktyg och kan vara lämpliga vid olika situationer. Det är till exempel fördelaktigt att likt Susskinds (2019) modell strukturera beslutsträd för enkla och binära beslut, medan mer komplex dataanalys kräver en självlärd algoritm.

Resultatet och det teoretiska ramverket visar att det inte går att skapa en enhetlig lösning, utan att ett AI-verktyg måste utvecklas utifrån vilket syfte den ska användas. Domstolarna måste specificera vilken typ av uppgift som ska automatiseras och sedan utifrån AI:ns tilltänkta uppgift ställa krav på förklarbarhet, transparens och prestanda. Till exempel skulle en viss typ av uppgift kunna lösas bäst med hjälp av ett neuralt nätverk. Denna uppgift ställer dock, på grund av sin natur, även höga krav på modellens transparens och förmåga att förklara sina resultat, något som bayesiska nätverk istället är bra på. I ett sådant fall måste en avvägning av vad som är viktigast göras; modellens förmåga att mer konsekvent generera korrekta slutsatser eller dess förmåga att förklara dem.

Sverige skulle i utvecklingen av AI för domstolsväsendet med fördel kunna iaktta och dra nytta av andra länders tekniska kompetenser och praktiska erfarenheter. Vid en lyckad implementering finns dessutom möjligheten att exportera ett AI-verktyg till länder med lägre grad av teknisk utveckling. Detta knyter an till ett av FN:s hållbarhetsmål som rör bland annat utbyten av kunskap, teknik och expertis.

5.4 Tillvägagångssätt vid implementering

Det finns många åsikter kring AI inom domstolsväsendet, inom vilka områden en implementering skulle kunna vara aktuell, hur en sådan implementering i praktiken skulle gå till och vilka konsekvenser den skulle kunna få. I slutändan är det ingen som säkert vet hur en implementering skulle påverka domstolsväsendet. Denna kunskap kan endast fås genom att initiera småskaliga tester och se vilka effekter AI har på domstolarna. Det skulle till exempel vara möjligt för en tingsrätt att testa en implementering av någon form av AI-verktyg vars uppgift har mycket begränsad mänsklig påverkan. Visar sig denna implementering vara framgångsrik kan verktyget börja användas mer utbrett bland svenska domstolar. Det kan dock inte bara släppas fritt, utan det måste ske kontinuerlig monitorering av dess funktion och beteende.

Någorlunda okontroversiella områden för implementering inom det svenska domstolsväsendet inkluderar dokumenthantering i allmänhet och översättning samt arkivering i synnerhet. Detta eftersom resultatet av testning inom dessa områden sannolikt inte skulle ha någon större mänsklig påverkan. Inom dessa områden skulle dagens AI ha goda

förutsättningar att effektivisera processer, men även förbättra processernas kvalitet om implementeringen sker på rätt sätt. Dessutom har sådan AI utvecklats mycket bara de senaste åren, fortsätter det i samma riktning blir det med tiden svårmotiverat att inte adoptera tekniken.

I dagsläget skulle AI också kunna assistera i fall som mer eller mindre är att betrakta som helt binära. Som tidigare nämnts i diskussionen är ett exempel på detta skilsmässor, mer specifikt gällande beviljandet av betänketid. Just detta är en standardiserad process som bör ha goda förutsättningar att automatiseras. Däremot bör inte AI-verktyget ha ansvarsbördan för de beslut som fattas, den ska fortsatt ligga på mänsklig nivå.

En självlärd AI som skulle ha en mer direkt påverkan på utfall av mål är idag relativt avlägsen. Testning av en sådan AI bör föregås av att enklare AI-verktyg implementeras för att skapa nödvändig kunskap och förståelse kring AI. Det är heller inte sannolikt att tekniken är tillräckligt mogen för att kunna appliceras i en rådgivande funktion och samtidigt bidra till en faktisk förbättring. En framtida implementering inom detta område förutsätter ett gediget arbete i att förhindra oönskade bias i algoritmen och dataseten på vilka den tränas.

Oavsett vilken typ av AI-verktyg som testas måste det utvecklas med ett tydligt syfte utifrån det problem som ska lösas. Möjligheten att skapa en generell AI som kan lösa flera uppgifter finns inte, dagens tekniska förutsättningar kräver specifika lösningar. För olika typer av mål och problem finns olika kravbild på exempelvis transparens, förklarbarhet samt datas kvantitet och kvalitet. Därmed finns det ingen enskild lösning som går att applicera överallt, det varierar helt beroende på problem och situation.

5.5 Vidare forskning och utredning

Under studiens gång har flertalet områden identifierats där det behöver utföras vidare forskning för att få en klarare bild över hur en framgångsrik AI-implementering skulle se ut. Ett sådant område är forskningen kring urval för bias, där det idag finns mycket kvar att studera rörande urvalet av gamla fall. Detta är i dagsläget ett relativt kontroversiellt fält, då det är upp till en människa att välja ut vilka gamla domar som ska ingå i datasetet.

Som tidigare nämnts i diskussionen är ansvarsfrågan ett annat område där det finns potential för vidare utredning, hur ansvaret ska fördelas, speciellt med tanke på att användning av AI ökar mer och mer, och att AI idag har en allt större samhällspåverkan. Hur mycket ansvar ska läggas på användare respektive utvecklare? Ska ansvaret för dessa implementeringar ligga på domstolarna själva, eller ska det föras upp på statlig nivå? Måste en domare bära ansvarsbördan för ett beslut fattat av en AI, som denne kanske inte ens står bakom? I ansvarsfrågan finns det många frågor som behöver utredas vidare, i syfte att skapa underlag och regelverk som underlättar framtida implementeringar.

6. Slutsats och rekommendation

I denna rapport har möjligheten att tillämpa artificiell intelligens inom domstolsväsendet utforskats. I syfte att utreda hur en sådan implementering skulle kunna se ut och vad som skulle vara de huvudsakliga möjligheterna, riskerna och begränsningarna har intervjuer genomförts med experter inom såväl AI som juridik. Dessutom har ett teoretiskt ramverk skapats med avsikt att konstruera ett underlag för diskussion.

Studiens resultat påvisar att en implementering av AI inom det svenska domstolsväsendet i viss mån är möjlig. AI skulle idag kunna nyttjas inom flera områden, som bland annat arkivering, översättning och annan dokumenthantering. Långsiktigt kan AI med en beslutsstödjande funktion vara aktuellt, men i dagsläget saknas både tekniska och samhälleliga förutsättningar. Idag går det inte att förutspå hur tekniken och samhället kommer att utvecklas samt vilka möjligheter och begränsningar som det medför. Därför ska det inte uteslutas att AI skulle kunna ha en central roll i framtidens domstolar.

Effekten av AI i det svenska domstolsväsendet bör enligt resultatet bli mer resurseffektiva domstolar, dels eftersom processtiden kortas, dels för att mänskliga resurser kan omfördelas. Andra potentiella fördelar med att nyttja AI inom domstolsväsendet är att domstolarna blir mer enhetliga, samt att kvaliteten på dömandet ökar. Detta kan i sin tur leda till ökad rättssäkerhet.

I studien har det framkommit att det finns flera aspekter att ta i beaktning vid implementering av AI. Algoritmens output behöver motiveras och förklaras tydligt, samtidigt som en hög förklarbarhet kan bidra till att öka allmänhetens förtroende för verktyget. Huruvida källkoden behöver vara en offentlig handling och verktyget därmed fullständigt transparent råder det dock delade meningar om. Däremot ska det finnas möjlighet att kontinuerligt granska och kvalitetssäkra koden. Det är också av stor vikt att utvecklare och användare är medvetna om biasproblematiken, och att de säkerställer att algoritmen inte präglas av oönskade biaser.

Studien visar att det inte går att utveckla en generell AI-lösning för domstolar, utan att den måste utvecklas utifrån ett specifikt syfte. För olika problem lämpar sig olika metoder och modeller, med varierande grad av komplexitet och förklarbarhet.

Resultatet av denna undersökning visar också att innan AI kan implementeras i en skarpare miljö krävs det att ansvarsfördelningen mellan utvecklare och användare utreds, samt att prestandakrav för AI:n preciseras. Att internt testa olika typer av AI-verktyg i liten skala, vilket eventuellt skulle kunna ske parallellt med nuvarande tillvägagångssätt, behöver dock inte vara omöjligt fastän ovanstående frågor inte är helt klarlagda. Sådan testning skulle potentiellt bidra till att öka kunskapen och kompetensen inom området.

Resultatet och erfarenheter från skapandet av det teoretiska ramverket visar att det på många håll saknas forskningsunderlag gällande kombinationen AI och juridik. Det krävs i allmänhet mer forskning kring området, specifikt i bias- och transparensfrågor. Praktisk testning i samband med eller som komplement till forskning kan vara ett lämpligt tillvägagångssätt för att effektivt bygga upp kompetens och erfarenhet.

Sammanfattningsvis har studien identifierat fyra särskilt viktiga punkter att ta i beaktning i fortsatt forskning och innan påbörjad testning:

- Se till att AI:n utvecklas utifrån ett tydligt syfte.
- Tydliggör ansvarsfördelningen.
- Säkerställ att AI:n utvecklas med förklarbarhet och transparens i fokus.
- Var medveten om och försöka minimera oönskad bias, både hos människor och dataset.

Källförteckning

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). *Machine Bias*. ProPublica. Hämtad 2020-03-26 från https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Baer, T. (2019). *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists* (1:a upplagan). New York: Apress Media.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (1:a upplagan). Singapore: Springer.

Blomkvist, P. (2014). *Metod för teknologer: Examensarbete enligt 4-fasmodellen* (1:a upplagan). Lund: Studentlitteratur.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies* (1:a upplagan). Oxford: Oxford University Press.

Brottsförebyggande rådet. (2017). *Enhetligt dömande i tingsrätter: En statistisk analys av andelen fängelsedomar*. (URN: NBN: SE: BRA-725). Hämtad 2020-02-04 från https://www.bra.se/publikationer/arkiv/publikationer/2017-09-01-enhetligt-domande-i-tingsratter.html

Bryman, A. & Bell, E. (2015). *Business research methods* (4:e utgåvan). Oxford: Oxford University Press.

Cambridge Dictionary. (u.å.). *Bias*. Hämtad 2020-05-07 från https://dictionary.cambridge.org/dictionary/english/bias

Corbett-Davies, S. & Goel, S. (2018). *The measure and mismeasure of fairness: A critical review of fair machine learning*. Hämtad 2020-03-20 från https://arxiv.org/abs/1808.00023

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. & Aziz Huq. (2017). Algorithmic Decision Making and the Cost of Fairness. *In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. 797–806. DOI: https://doi.org/10.1145/3097983.3098095

Dietrich, W., Mendoza, C. & Brennan, T. (2016). *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity.* Hämtad 2020-03-31 från http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf

Dressel, J. & Farid, H. (2018). *The accuracy, fairness, and limits of predicting recidivism.* Science advances. *Vol. 4, no. 1.* (eaao5580). DOI: 10.1126/sciadv.aao5580.

Dutchnews.nl. (2020). Government's fraud algorithm SyRI breaks human rights, privacy law. *Dutchnews.nl.* Hämtad 2020-03-26 från https://www.dutchnews.nl/news/2020/02/governments-fraud-algorithm-syri-breaks-human-rights-privacy-law/

European Commission for the Efficiency of Justice. (2018). *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. Hämtad från https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment

Family Health International. (u.å.). *Qualitative Research Methods Overview*. Hämtad 2020-02-07 från https://course.ccs.neu.edu/is4800sp12/resources/qualmethods.pdf

Faulkner, S.L., & Trotter, S.P. (2017). Data Saturation. I *The International Encyclopedia of Communication Research Methods*. Hämtad från https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118901731.iecrm0060

Flores, A.W., Lowenkamp, C.T. & Bechtel, K. (2016). *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to*

Predict Future Criminals. And it's Biased Against Blacks". Hämtad 2020-04-01 från http://www.crj.org/assets/2017/07/9_Machine_bias_rejoinder.pdf

Future of Life Institute. (2017). *Asilomar AI principles* Hämtad 2020-02-21 från https://futureoflife.org/ai-principles/

Förenta Nationerna. (2008) *Allmän förklaring om de mänskliga rättigheterna.* Hämtad från https://fn.se/wp-content/uploads/2016/07/Allmanforklaringomdemanskligarattigheterna.pdf

Förvaltningsrätten i Göteborg. (2019). *Handläggningstider*. Hämtad 2020-02-10 från https://www.domstol.se/forvaltningsratten-i-goteborg/om-forvaltningsratten/handlaggningstider/handlaggningstider/

Förvaltningsrätten i Linköping. (2019). *Handläggningstider*. Hämtad 2020-02-10 från https://www.domstol.se/forvaltningsratten-i-linkoping/om-forvaltningsratten/aktuellt/handlaggningstider/

Förvaltningsrätten i Malmö. (2019). *Handläggningstider*. Hämtad 2020-02-10 från https://www.domstol.se/forvaltningsratten-i-malmo/om-forvaltningsratten/organisation/handlaggningstider/

Geiger, R.S., Yu, K.,Yang, Y., Dai, M., Qiu, J., Tang, R., & Huang, J. (2019). Garbage In, Garbage Out? Do Machine Learning Application Papers in Social Computing Report Where Human-Labeled Training Data Comes From? *Proc ACM FAT* 2020.* doi:10.1145/3351095.3372862.

Griffiths, P., Gossop, M., Powis, B., & Strang, J. (1993). Reaching hidden populations of drug users by privileged access interviewers: methodological and practical issues, *Addiction*, vol. 88, 1617-1626.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods*. 18(1) 59-82. doi: 10.1177/1525822X05279903.

High Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI.* Hämtad 2020-04-16 från https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Häggström, O. (2019). Artificiell Intelligens, Stora Möjligheter och Stora Risker. *Fysikaktuellt*, (1), 14-18.

Jordan, M.I. (2014). Neural Networks. I Tucker (Red.), *Computing Handbook.* (3:e upplagan). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Knutas, J., & Bergström, A. (2018). *"..att en blir sittande med ett hjärndött administrativt jobb": En kvalitativ studie om digitaliseringen och effektiviseringen inom ekonomiskt bistånd.* (Examensarbete, Linnéuniversitetet, Kalmar/Växjö). Hämtad från http://lnu.diva-portal.org/smash/get/diva2:1215943/FULLTEXT01.pdf

Lag (SFS 2010:1408). Hämtad från Riksdagens webbplats: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/kungorelse-1974152-om-beslutad-ny-regeringsform_sfs-1974-152

Lantz, A. (2013) *Intervjumetodik* (Upplaga 3:1). Lund: Studentlitteratur.

MacMillen, A. (2019) *Can an Algorithm Identify Repeat Offenders?* Hämtad 2020-03-26 från https://chicagopolicyreview.org/2019/03/12/can-an-algorithm-identify-repeat-offenders/

Mahanta, J. (2017) *Introduction to Neural Networks, Advantages and Applications.* Hämtad 2020-03-17 från https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207

Marsland, S. (2014). Machine Learning. I Tucker (Red.), *Computing Handbook.* (3:e upplagan). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Mitchell, T.M. (1997) *Machine learning* (1:a upplagan). New York, NY: McGraw-Hill.

Mueller, E.T. (2015). *Commonsense Reasoning*. (2:a upplagan). Waltham, MA: Elsevier Inc.

Myndigheten för digital förvaltning. (2019). *Främja den offentliga förvaltningens förmåga att använda AI. (*I2019/01416/DF). Hämtad från https://www.digg.se/globalassets/slutrapport---framja-den-offentliga-forvaltningens-formaga-att-anvanda-ai.pdf

Artificiell intelligens. (2020). I *Nationalencyklopedin.* Hämtad från *http://www.ne.se/uppslagsverk/encyklopedi/lång/artificiell-intelligens*

Nielsen, M. (2019). *Using neural nets to recognize handwritten digits*. Hämtad 2020-03-17 från: http://neuralnetworksanddeeplearning.com/chap1.html

Nilsson, N. (1965). *Learning Machines: Foundations of trainable pattern-classifying Systems.* (1:a upplagan). New York, NY: McGraw-Hill.

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, *11*(4), 27-344. doi: 10.1080/13645570701401305.

Osborne, C. (2020) Dutch court rules AI benefits fraud detection system violates EU human rights. *Zero Day*. Hämtad 2020-03-26 från https://www.zdnet.com/article/dutch-court-rules-ai-benefits-fraud-detection-system-violates-eu-human-rights/

Pearl, J. (2014). Graphical Models for Probalistic and Causal Reasoning. I Tucker (Red.), *Computing Handbook* (3:e upplagan., kap. 42) Boca Raton, FL: CRC Press, Taylor & Francis Group.

Regeringen. (2018). *Regeringens proposition 2018/19:1: Budgetpropositionen för 2019 Förslag till statens budget för 2019, finansplan och skattefrågor*. Hämtad från https://www.regeringen.se/4ac33b/contentassets/d13d35490a9f470a87b885188587b5ae/budgetpropositionen-for-2019-hela-dokumentet-prop.-2018191.pdf

Regeringskansliet (2019). *Satsningar inom Justitiedepartementets ansvarsområden i budgeten.* Hämtad 2020-03-16 från https://www.regeringen.se/pressmeddelanden/2019/09/satsningar-inom-justitiedepartementets-ansvarsomraden-i-budgeten/

Rissland, E. L., Ashley, K. D., & Loui, R. P. (2003). AI and Law: A fruitful synergy. *Artificial Intelligence*, 150(1-2), 1–15.

Scaramuzzino, G. (2019). *Socialarbetare om automatisering i socialt arbete: En webbenkätundersökning.* Hämtad 2020-03-26 från http://lup.lub.lu.se/search/ws/files/61894787/RRSW_2019_3.pdf

Susskind, R. (2019). *Online Courts and the Future of Justice.* Oxford: Oxford University Press.

Svensson, L., & Larsson, S. (2017). *Digitalisering och socialt arbete - en kunskapsöversikt.* Hämtad 2020-03-26 från *https://lup.lub.lu.se/search/ws/files/22297951/Svensson_Larsson_2017_Digitalisering_och_socialt_arbete_en_kunskaps_versikt.pdf*

Svensson, L. (2019). *"Tekniken är den enkla biten": Om att implementera digital automatisering i handläggningen av försörjningsstöd.* RESEARCH REPORTS IN SOCIAL WORK: Socialhögskolan, Lunds universitet.

Stanford University. (u.å.). *Professor John McCarthy*. Hämtad 2020-05-10 från http://jmc.stanford.edu/index.html

Trelleborgs kommun. (2019). *Så här hanteras en digital ansökan.* Hämtad 2020-02-09 från https://www.trelleborg.se/sv/omsorg-hjalp/ekonomiskt-stod/ekonomiskt-bistand-forsorjningsstod/sa-har-hanteras-en-digital-ansokan/

UNDP. (2015). Om globala målen. Hämtad 2020-05-03 från https://www.globalamalen.se/om-globala-malen/

United Nations Human Rights. (2019). The Netherlands is building a surveillance state for the poor, says UN rights expert. Hämtad 2020-03-26 från https://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=25152&LangID=E

Wallén, G. (1996). *Vetenskapsteori och forskningsmetodik* (Upplaga 2:15). Lund: Studentlitteratur.

Wired. (2018) When it comes to gorillas, Google Photos remains blind. Hämtad: 2020-04-02 från https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

Bilagor

Intervjumall

1. Frågor till respondenter inom området juridik

- Är du okej med att vi spelar in intervjun?

- Hur får vi använda ditt namn i rapporten, om anonym gäller det bara direkta citat eller vill du inte att ditt namn ska förekomma?

- Kan du berätta lite om dig själv, bakgrund, utbildning, nuvarande befattning?

- Vad är det första du tänker på du när du hör orden artificiell intelligens?

- Har du någon erfarenhet av AI inom rättsväsendet? Antingen stött på det eller hört/läst om det?

- Vilka fördelar, nackdelar respektive begränsningar ser du med AI?

- Inom vilka områden av domstolsväsendet skulle AI kunna bidra mest, enligt dig?

- Till vilken grad skulle det kunna implementeras tror du?
    - Från beslutsunderlag till eget dömande.
    - I vilka brottmål kan det vara aktuellt? Var drar man gränsen, rattfylla? Mord?

- Ur ett etiskt perspektiv, vad är viktigt att tänka på vid en implementering av AI?

- Vem bär ansvar för domar eller beslut fattade av en AI? Hur ska det hanteras?

- Hur bör transparensproblematiken hanteras?

- Vad tycker du om att implementera AI i svenska domstolar?

- Skulle du själv kunna tänka dig att i ett brott- eller tvistemål bli helt eller delvis dömd av en AI?
- Finns det någonting uppenbart som vi har missat? Vad tycker du att vi borde ha frågat dig om?

2. Frågor till respondenter inom området AI

- Är du okej med att vi spelar in intervjun?

- Hur får vi använda ditt namn i rapporten, om anonym gäller det bara direkta citat eller vill du inte att ditt namn ska förekomma?

- Kan du berätta lite om dig själv, bakgrund, utbildning, nuvarande befattning?

- Vad är det första du tänker på du när du hör orden Artificiell Intelligens?

- Hur ser du på implementering av AI i samhället?
    - Fördelar respektive nackdelar?
    - Mer om domstolar specifikt?

- Ur ett etiskt perspektiv, vad är viktigt att tänka på vid implementering av AI?

- Till vilken nivå är det rimligt att tro att man kan implementera AI?
    - Från beslutsunderlag till eget dömande.
    - I vilka brottmål kan det vara aktuellt? Var drar man gränsen, rattfylla? Mord?

- Finns det någon risk att det uppstår problematik med “black box” och transparens?
    - Hur skulle man gå tillväga för att hantera detta?

- Vem bär ansvar för domar eller beslut fattade av en AI? Hur ska det hanteras?

- Vad tycker du om att implementera AI i svenska domstolar?

- Skulle du själv kunna tänka dig att i ett brott- eller tvistemål bli helt eller delvis dömd av en AI?

- Finns det någonting uppenbart som vi har missat? Vad tycker du att vi borde frågat dig om?

INSTITUTIONEN FÖR TEKNIKENS EKONOMI OCH ORGANISATION
AVDELNINGEN FÖR ENTREPRENÖRSKAP OCH STRATEGI
CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2020
www.chalmers.se

CHALMERS