HÖGSKOLAN
I GÄVLE

AKADEMIN FÖR TEKNIK OCH MILJÖ
Avdelningen för industriell utveckling, IT och samhällsbyggnad

Artificiell intelligens som beslutsmetod

Micael Frideros

2017

Examensarbete, Avancerad nivå (magisterexamen), 15 hp
Besluts-, risk- och policyanalys
Magisterprogram i besluts-, risk- och policyanalys

Handledare: Magnus Hjelmblom
Examinator: Fredrik Bökman

Artificiell intelligens som beslutsmetod

av

Micael Frideros

Akademin för teknik och miljö
Högskolan i Gävle

S-801 76 Gävle, Sweden

Email:
micael.frideros@gmail.com

Abstrakt

Detta arbete behandlar artificiell intelligens som beslutsmetod. Efter inledande diskussioner om de övergripande skillnaderna mellan hjärnans och datorns funktionssätt, olika utvecklingsinriktningar av artificiell intelligens samt olika metoder för att skapa artificiell intelligens identifieras strategier för hur artificiell intelligens kan användas som beslutsmetod beroende på faktorer som transparens, effektivitet samt mängden tillgänglig testdata. Exempelvis identifieras några typiska beslutssituationer där det kan antas att automatiserat beslutsfattande baserat på artificiell intelligens har stor potential, samt situationer då metoden kan antas vara mindre lämplig. Därefter analyseras teknikutvecklingen inom artificiell intelligens, både generellt och inom fyra specifika tillämpningsområden: inom autonoma fordon, inom finans, inom medicin och inom militären. Både den övergripande undersökningen av den generella teknikutvecklingen och studien av de fyra teknikområdena indikerar fortsatt mycket snabb utveckling inom området. Exempelvis visar en analys av patentdatabasen Espacenet att antalet patent inom området utvecklas i det närmaste exponentiellt. Samtidigt har det på senare tid gjorts flera tekniska genombrott, t.ex. utvecklandet av allt effektivare algoritmer genom användandet av hierarkiska strukturer med flera olika nivåer av ickelinjär informationsbearbetning, något som ofta benämns Deep Learning. Ett exempel är den metod för artificiell intelligens som utvecklas av DeepMind, som visat sig vara tillämpningsbar inom många olika områden, från att spela klassiska datorspel som Space Invaders och Breakout på en övermänsklig nivå till att göra betydande effektiviseringar i driften av Googles datorhallar. Även ur ett hårdvaruperspektiv är utvecklingen närmast exponentiell, driven av kontinuerliga framsteg inom tillverkningsprocesser samtidigt som det nyligen gjorts betydande framsteg med specialiserade kretsar för artificiell intelligens, något som sannolikt kommer att resultera i ännu snabbare utveckling av kraftfullare artificiell intelligens inom en nära framtid. Med hänsyn till teknikens effektivitet och den mycket snabba utvecklingen inom området diskuteras även några specifika frågeställningar som ofta nämns i diskussionen om artificiell intelligens, dess påverkan på arbetsmarknaden och den globala säkerhetsbalansen, för att baserat på detta sedan diskutera artificiell intelligens som beslutsmetod även i ett vidare perspektiv.

Innehåll

1 Inledning

Evolutionen har gett oss människor en fantastisk hjärna, kapabel till bedrifter som saknar motstycke på jorden. Vi har tämjt elden, satt hjulet i rullning och utvecklat ett rikt språk. Genom språket kan vi resonera, kommunicera och samarbeta med varandra, och genom skrivtecken har vi ökat språkets räckvidd över såväl tid som rum.

Genom språk och samarbete har vi tillsammans exempelvis kunnat utveckla kunskap om mediciner, vacciner, mekanik, gravitation, byggnadsteknik och fysik. Vi har bemästrat atomen både som energikälla och som vapen, och vi har lärt oss använda elektronen för såväl energiöverföring som informationsbärare. Med elektronen som informationsbärare har vi utvecklat tekniker som transistorer och kiselbaserade datorkretsar, tekniker som bland annat tillämpats i radioapparater, telefoner och tv, och i kombination med kunskaper inom abstrakt matematik även i datorer, smartphones och surfplattor.

Den tekniska utvecklingen av effektivare och kraftfullare datorkretsar har gått fort och tycks gå allt fortare. Intels grundare Gordon E Moore förutspådde 1965 att den tekniska utvecklingen av allt mindre kretsar skulle medföra att antalet transistorer på ett datorchip, och därmed beräkningskraften, skulle fördubblas vartannat år – en uppskattning som hittills visat sig stämma bra (Intel, 2016).

Parallellt med att datorerna har blivit allt kraftfullare så har man även utvecklat teknik och metoder för att göra dem intelligentare genom att de fått metoder att själva lära sig och dra slutsatser, ofta kallat maskininlärning (eng machine learning). Metoder som exempelvis låter programmet självt lära sig en optimal strategi och anpassa sig efter nya förutsättningar. Datorbaserat beslutsfattande med utgångspunkt i olika algoritmer har på många sätt blivit en del av vår vardag, även om vi sannolikt oftast inte tänker på det. Googles sökningar, Amazons annonser, Spotifys spellistor och självgående dammsugare styrs alla av algoritmer, ofta med en förmåga att kontinuerligt utvecklas och anpassa sig till förändringar.

På senare år har datorerna blivit så kraftfulla att de kan användas för program som de slår de bästa spelarna i schack, Jeopardy och go. Men denna typ av teknik har även blivit allt vanligare inom exempelvis aktie- och valutahandel, där datorprogram på millisekunder fattar beslut om att köpa eller sälja värdepapper. Exempelvis stod algoritmhandeln för 45 % och högfrekvenshandeln för 13 % av all handel på stockholmsbörsen 2011 enligt Nasdaq OMX Stockholm (Sveriges Riksbank, 2015, s.56). Tekniken börjar även användas alltmer inom exempelvis trafiken, medicinen och militären, tillämpningar som väcker frågor om tekniken som beslutsmetod.

1.1 Begreppsdiskussion

I denna uppsats behandlas flera olika uttryck som för tydlighetes skull kan diskuteras och definieras. Artificiell intelligens är en sammansatt term av artificiell (d.v.s. konstgjord) och intelligens. Ordet intelligens härstammar av latinets *intellego*, som kan översättas med att förstå, begripa, inse, avgöra. På grund av de kognitiva begränsningarna i dagens datasystem är det idag knappast särskilt meningsfullt att i denna kontext diskutera AI-systemens förståelse eller förmåga att begripa och inse. Däremot har AI-system i någon mening förmåga att anpassa sitt beteende efter den aktuella situationen och när man talar om artificiell intelligens menar jag därför att det är just denna förmåga som är central.

Bland AI-systemen finns en kategori som är utrustade med mekanismer för inlärning, sk machine learning, som gör dem mer flexibla att utveckla sitt beteende bortom förprogrammerade handlingsmönster och kan därför anpassa sitt beteende till nya förutsättningar. Dessa tillämpningar kan därför sägas vara kraftfullare och i någon mening intelligentare än de som bara följer fördefinierade handlingsmönster och detta arbete kommer därför framförallt fokusera på dessa metoder och tillämpningar.

I arbetet används även begrepp som beslutsmetod och beslutssystem. Med beslutsmetod avses ett tillvägagångssätt eller en teknik som kan användas för att fatta beslut, medan beslutssystem avser system som i någon mening kan användas vid beslutsfattande, antingen för att självständigt fatta beslut på egen hand eller genom att på olika sätt stödja den mänskliga beslutsprocessen.

1.2 Syftesdiskussion

Det finns många intressanta och viktiga frågeställningar att studera och reflektera kring beslutsfattande genom självlärande beslutsalgoritmer, sk artificiell intelligens, både ur ett individuellt och samhälleligt perspektiv. Många har studerat hur man bäst kan använda olika typer av algoritmer vid olika beslutssituationer. Vad är det exempelvis i en tillämpning som avgör om en viss typ av algoritm passar bättre än någon annan och hur kan man på bästa sätt anpassa maskininlärningen så att algoritmen blir så träffsäker som möjligt även i nya situationer?

Andra har intresserat sig för olika säkerhetsaspekter kring denna typ av beslutssystem. Hur kan man exempelvis veta att besluten som denna typ av system fattar alltid är i linje med de ursprungliga intentionerna och hur kan man försäkra sig om att systemen inte manipuleras av någon utomstående?

Ytterligare några har funderat mycket över vart utvecklingen med artificiell intelligens kan tänkas leda, i takt med att vi utvecklar allt starkare artificiell intelligens. Vad innebär det om vi en dag utvecklar en artificiell intelligens som är mer kraftfull än människans hjärna – och denna intelligens sedan används för att utveckla tekniken ännu längre?

En intressant aspekt är att tekniken med artificiell intelligens kan sägas vara en ”omstörtande innovation” (jfr eng disruptive innovation), men inte bara gällande en viss produkt eller marknad, utan för världsekonomin i stort (Manyika *et al*., 2013, s.70). Vad händer i samhället i stort när mänskligt beslutsfattande ersätts av artificiell intelligens – i arbetslivet och ekonomin, inom vård och omsorg, inom det militära?

1.3 Syfte

Detta arbetes syfte är att diskutera och analysera artificiell intelligens som beslutsmetod. Genom att diskutera funktionsskillnader mellan den mänskliga hjärnan och datorer identifieras situationer när artificiell intelligens kan antas ha ett övertag jämfört med mänskligt beslutsfattande. Därefter diskuteras även olika metoder för artificiell intelligens ur ett beslutsperspektiv och några faktorer som man bör ta hänsyn till vid val av metod.

Med tanke på teknikens kraftfullhet, den mycket snabba utvecklingen inom området och den intensiva debatten om teknikens utveckling och tillämpningar menar jag att en diskussion om artificiell intelligens som beslutsmetod är ofullständig utan att samtidigt reflektera över de konsekvenser som en ökad användning av tekniken kan antas medföra. Därför avser jag också att undersöka utvecklingen av artificiell intelligens och delar av den internationella debatten inom området för att efter en analys av dessa komma fram till några övergripande slutsatser om utvecklingen och dess sannolika konsekvenser.

2 Visionen om bättre beslutsfattande genom artificiell intelligens

Samhällsutvecklingen i stort har bland annat medfört ökad kunskap och valfrihet, men detta har också medfört att beslutssituationerna blivit allt mer komplexa. Beslut som tidigare upplevdes som relativt enkla, exempelvis ”jag behöver handla mjölk och lite mat – ska jag välja Coop eller Ica”, har idag inte bara fått fler direkta alternativ som minimjölk, lättmjölk, mellanmjölk, standardmjölk, lantmjölk eller kanske laktosfri och var kan jag handla; ”Hemköp, Lidl, Vivo, Willys, Netto eller Tempo” utan även flera andra konkurrerande alternativ, som att exempelvis handla via internet eller beställa en matkasse. Eller varför inte beställa hemkörd mat från en restaurang, och ska jag i så fall äta thai, indiskt, kina, hamburgare, sushi, kebab eller kanske en pizza? Och sedan var det ju det där med klimatet, vad är mest klimatsmart – ta bilen till affären eller beställa hem, eller kanske använda en el-cykel – och vilken mat är mest klimatsmart? Sedan är det ju näringsinnehåll och nyttighet, och ekonomi, och global rättvisa, för att bara nämna några parametrar.

På samma sätt möter vi idag många andra mer komplexa och ofta även abstrakta beslutssituationer än tidigare, exempelvis relaterade till ekonomiska frågor som pension eller boende. Men när beslutssituationen uppfattas som allt mer komplex så blir beslutsfattandet allt mer krävande för oss. Inte minst när beslut ska fattas i situationer som upplevs som osäkra, exempelvis när vi ska fatta beslut som får konsekvenser långt in i en osäker framtid, när det finns många olika tänkbara perspektiv som kan leda till olika slutsatser eller när det finns många olika mål, som kanske står i motsatsförhållande till varandra – ”man kan ju inte både äta kakan och ha den kvar” kan exempelvis sägas vara en metafor för en vanlig målkonflikt. (Clemen & Reilley, 2001, s.2-3)

Ett annat problem med mänskligt beslutsfattande är att vår hjärna inte alltid fungerar så rationellt som vi skulle önska, något som alltför ofta tycks leda till misstag. Exempelvis har vi ofta svårt att värdera sannolikheter, eller göra nyttojämförelser när något är gratis. Ofta använder vi oss av tumregler eller heuristiker när vi fattar beslut, något som kan fungera bra men ibland kan leda fel (Kahneman, 2013, Ariley, 2008). I artikeln ”Judgement under uncertainty: Heuristics and Biases” konstaterar exempelvis Tversky & Kahneman (1974):

> In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

Ytterligare ett exempel på problem med mänskligt beslutsfattande är det mänskliga sinnets svårigheter att fatta rutinmässiga och återkommande beslut med konstant och hög kvalitet, något som exempelvis leder till olyckor. En studie vid Uppsala universitet fann exempelvis att ”den mänskliga faktorn” var inblandad i 90-95 % av alla trafikolyckor (Forward, 2008). Med tanke på att det enligt transportstyrelsens olycksstatistik dör mellan 250 och 400 personer i den svenska trafiken varje år, och mellan 17 000 och 26 000 skadas, så är det lätt att se den potentiella nyttan av säkrare självkörande bilar (Transportstyrelsen, 2016). Även inom flyget och sjöfarten har man utvecklat stödsystem, t.ex. autopilot, för att hantera denna mänskliga beslutssvaghet.

Med andra ord tycks det finnas två huvudsituationer där mänskligt beslutsfattande har påtagliga svagheter. Dels i mer komplexa beslutssituationer med stora datamaterial och mycket statistik, eller i situationer som är komplexa på grund av flera olika och kanske motstridiga mål, och dels att fatta rutinmässiga beslut med hög och jämn tillförlitlighet.

Som ett svar på svagheterna har det bland annat utvecklats tekniker för mer tillförlitligt beslutsfattande, exempelvis genom att använda olika strukturerade metoder för beslutsfattande eller genom att ta hjälp av datorer för exempelvis simuleringar och beräkningar. Samtidigt har det de senaste åren gjorts betydande framsteg inom lärande algoritmer – något som ofta kallas ”machine learning” eller

”artificiell intelligens”. En dator kan snabbt och effektivt bearbeta mycket stora mängder data med tillförlitligt resultat och blir inte trött utan håller en konstant kvalitetsnivå, något som lett till en vision att skapa en maskin som kan hjälpa människan genom att fatta bättre beslut.

2.1 Skillnader mellan hjärna och dator

Eftersom hjärnan är den intelligenta beslutsprocessor som vi alla använt sedan födseln och därför är vana vid att använda är den en naturlig utgångspunkt vid diskussioner om intelligens. Dock finns det flera centrala skillnader mellan hjärnans biologiska funktion och datorkretsar, som är viktiga att ta hänsyn till vid diskussioner om artificiell intelligens.

Exempelvis tar AI-forskaren Nilsson i sin bok ”The Quest for Artificial Intelligence” upp flera faktorer som han menar ofta nämns som fundamentala skillnader mellan hjärnan och datorn (Nilsson, 2010, s.392):

- Datorn har kanske något hundratal processorer medan hjärnan har ca 100 miljarder neuroner, med ca 100 biljoner kopplingar mellan sig (Cox, 2016).
- Datorer utför miljarder operationer per sekund medan hjärnan bara behöver utföra några tusen.
- Datorer går ner ibland medan hjärnan kan hantera felaktigheter utan att krascha.
- Datorer använder binära signaler medan hjärnan använder analoga.
- Datorer gör bara vad programmeraren beordrat den medan hjärnan kan vara kreativ och nyskapande.
- Datorer utför bara operationer i serie medan hjärnan fungerar parallellt.
- Datorer kan bara vara logiska medan hjärnan kan vara intuitiv.
- Datorer är programmerade medan hjärnor lär sig.

Som Nilsson själv konstaterar så är flera av dessa invändningar inte längre relevanta, exempelvis att en dator måste programmeras och inte kan lära sig. Samtidigt är andra fortfarande giltiga, exempelvis att datorer i någon mening måsta agera utifrån givna regler/instruktioner eller en uttalad målfunktion medan hjärnan kan vara både intuitiv och kreativ.

Denna funktionsskillnad beror sannolikt till stor del på skillnaden mellan den logik som den mekaniska strukturen ger upphov till. En datorprocessor är i grunden uppbyggd binärt kring ”ettor” och ”nollor”, något som utvidgas till fler betydelser genom sekvenser av ettor och nollor. Dessa kan exempelvis regelstyrt bearbetas, adderas eller subtraheras med varandra i enlighet med förprogrammerade instruktioner, vilket leder till aritmetisk logik.

Hjärnan är å andra sidan uppbyggd av nätverk av neuroner med kopplingar som bland annat styrs av Hebbs regel, som ofta sammanfattas i uttrycket:

> Neurons that fire together, wire together. Neurons that fire out of sync, lose their link (Munz *et al*., 2014).

Med andra ord tycks kopplingen mellan neuroner som avfyras inom en viss tidsperiod förstärkas, medan kopplingen mellan neuroner som avfyras tidsmässigt oberoende av varandra försvagas. Detta medför bland annat att hjärnan är mycket föränderlig eller plastisk, och hela tiden anpassas efter de signaler och intryck som den registrerar. Eller som Rodney Douglas uttrycker det (Nicolelis, 2011, s.22):

> The brain truly works like an orchestra, but a unique one, in which the music it produces can almost instantaneously modify the configuration of its players and instruments and self-compose a whole new melody from this process.

Denna mekanism är sannolikt central för vår förmåga att lära oss och leder till en associativ logik som sannolikt kan förklara såväl psykologiska mekanismer som exempelvis betingning (Pavlovs hundar) som psykologiska behandlingsmetoder som kognitiv beteendeterapi (KBT) och neurolingvistisk programmering (NLP).

Ur ett neurovetenskapligt eller psykologiskt perspektiv saknar jag dock flera strukturella eller funktionsmässiga aspekter i Nilssons uppställning över skillnader mellan hjärnan och datorer. En väsentlig skillnad är att medan en inaktiv processor är tyst så kan en levande hjärna snarare alltid beskrivas som pulserande. Tekniker som optogenitik och EEG visar att hjärnan alltid är i någon form av spontan aktivitet, även när den inte får någon yttre stimuli eller ens är vaken. En levande hjärna genererar därför kontinuerligt nya tankar och intryck även i avsaknad av yttre stimuli, medan en tyst hjärna är en död hjärna. (Yuste, 2015)

En annan relaterad skillnad i funktionssätt som inte heller bör glömmas bort och som påverkar beslutsfattandet är slumpmässigheten i neuronens avfyrande av aktionspotentialer. Till skillnad från datorns processor baserad på transistorer som alltid fungerar på samma sätt så har neuronernas avfyrande av signaler, sk aktionspotentialer, ett påtagligt inslag av slumpmässighet. Detta tycks bland annat ha avgörande betydelse vid exempelvis inlärning (Engel *et al*., 2015), och kan sannolikt även mycket väl vara en viktig del av förklaringen till människans kreativitet.

Vidare menar jag att Nilssons uppställning saknar hjärnans modulära uppbyggnad med flera specialiserade funktioner som tycks ha mycket specifika uppgifter; t.ex. hippocampus, som fungerar som en slags kopplingsstation vid bland annat visualisering och återskapande av minnen och cerebellum (lillhjärnan), som innehåller den över 50 % av hjärnans neuroner trots att den bara står för ca 10 % av hjärnans volym. Cerebellums funktion är fortfarande ett intensivt forskningsområde, men anses allmänt ha en avgörande funktion för att reglerar rörelse och tycks sannolikt även ha en betydelsefull roll vid flera kognitiva uppgifter som exempelvis språkbehandling. (Cox, 2015a, Cox, 2015b)

Andra specifika strukturer som bör nämnas i ett beslutssammanhang är de som används för att värdera risk och nytta. Bland dessa finns exempelvis amygdala, som är starkt kopplad till både positiva och negativa känsloupplevelser, däribland upplevelsen av risk, fara och hot samt ventral striatum med nucleus accumbens, som ingår i hjärnans belöningssystem och som bland annat tycks vara aktiverat vid både upplevensen av nytta och värde i nuet och när vi beräknar framtida nytta och värde (Eagleman, 2015, Purves *et al*., 2012, Purves *et al*., 2013).

Med andra ord har den biologiska/mänskliga hjärnan flera specifika parallella neurala system, bland annat för funktioner som är centrala för upplevelsen av känslor, närvaro och medvetenhet. Man kan argumentera för att just kopplingen mellan belöningssystemen och inlärning är så centrala att de har varit avgörande för utvecklingen från enklare levande organismer till individer med upplevelser och medvetande (Ginsburg & Jablonska, 2007), något att både inspireras och kanske förskräckas över vid utveckling av artificiell intelligens.

Ytterligare en faktor som jag saknar i Nilssons uppställning är spegelneuroner, dvs neuroner som aktiverar motsvarande nervimpulser i vår hjärna som de som vi ser någon annan uppleva. Spegelneuronerna gör bland annat så att vi blir glada om vi ser någon som blir glad, om vi ser någon slå sig så gör det ont i oss etc. och är sannolikt en viktig förklaring till exempelvis egenskaper som empati, något som skiljer hjärnan från datorer och kan kanske ge värdefull inspiration till hur man exempelvis skulle kunna implementera moraliskt och empatiskt beteende hos AI (Nilzén, 2008).

Vidare är energiförsörjningen en relevant faktor när vi behandlar skillnader mellan datorer och hjärnor. Medan datorn får sin energi från en ständig ström av elektroner, så drivs hjärnans neuroner framförallt av den sk natriumkalium-pumpen, som skapar en obalans mellan natrium och kalium så att neuronen har ett överskott av kaliumjoner och ett underskott av natriumjoner jämfört med omgivningen. Eftersom naturen hela tiden strävar efter jämvikt kan denna obalans mycket snabbt korrigeras genom att

särskilda jonkanaler för kaliumjoner och natriumjoner öppnas, vilket leder till ett stort flöde av laddade atomer (joner) – dvs en elektrisk ström i form av en aktionspotential. Natriumkalium-pumpen och andra liknande processer är mycket energikrävande, något som exempelvis leder till att vi upplever mentalt arbete som beslutsfattande som krävande och uttröttande.

Energieffektivitet är sannolikt en stor del av förklaringen till de två tankesystem som psykologen Daniel Kahneman beskriver i sin bok ”Tänka, snabbt och långsamt”, som han döpt till system ett och system två. System ett är snabbt, automatiskt och utan känsla av medveten styrning medan system två är uppmärksamhetskrävande med känsla av kontroll, medvetna val och koncentration (Kahneman, 2013). Med tanke på hur energikrävande mental verksamhet är för hjärnan framstår det som rationellt att uppgifter som upplevs som enkla eller rutinmässiga sköts med minimal arbetsinsatts av tankesystem ett, medan tankesystem två får hantera mer krävande uppgifter som aktiva och medvetna beslut.

På ett mer övergripande plan menar jag att självmedvetenhet, eller medvetenhet om den egna existensen, är en annan viktig faktor att reflektera över i en diskussion om funktionsskillnader mellan datorer och biologiska hjärnor. Descartes (1637, s.22) ”je pense, donc je suis”, eller ”Cogito, ergo sum” på latin, är ett exempel på hur vi härleder vår existens ur mental närvaro, något som ai-kretsar saknar - hittills i alla fall.

Sammanfattningsvis kan det konstateras att det finns fundamentala skillnader mellan elektroniska datorkretsar och den biologiska hjärnan såväl gällande den mekaniska funktionen, exempelvis när det gäller tillförlitlighet och snabbhet, som den övergripande strukturen och när det gäller energiförsörjningen. Beteendemässigt leder detta bland annat till att datorkretsar beter sig regelstyrt, vilket å ena sidan ökar förutsägbarheten samtidigt som kretsarnas förmåga att tänka nytt och kreativt hindras. En annan påtaglig skillnad mellan datorkretsar och den mänskliga hjärnan är att datorn i sig själv saknar funktioner för självmedvetenhet och motsvarigheter till spegel-neuroner, något som ger möjlighet till både egoism och empati. Dessutom saknar datorer i sig själva de specifika kretsar som hjärnan har för värdering av nytta (ventral striatum) och risk (amygdala), något som vi människor ständigt analyserar och är medvetna om i form av känslor.

Dessa funktionella och beteendemässiga skillnader leder till både fördelar och nackdelar för artificiell intelligens i relation till mänskligt beslutsfattande. Bland de tillämpningar där artificiell intelligens kan ha påtagliga fördelar är exempelvis vid snabba, repetitiva och slentrianmässiga beslut, eller vid analys av stora datamängder. På samma sätt kan man tänka sig nackdelar vid beslut som kräver empati, etiska avvägningar eller kreativitet, eftersom vi inte lärt oss hur vi ska kunna återskapa dessa förmågor i en dator på ett sätt som motsvarar den mänskliga hjärnans kapacitet och funktionssätt.

2.2 Födelsen av artificiell intelligens

Tanken på tänkande, självlärande maskiner med en generell intelligens är på intet sätt ny, utan har diskuterats länge. Ett tidigt exempel är Alan Turings artikel ”Computing Machinery and Intelligence” från 1950, där han diskuterar processen med att skapa en tänkande och intelligent maskin med en intellektuell kapacitet liknande en vuxen människas genom att ta inspiration från ett barns utveckling (Turing, 1950, s.452):

> In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components.
>
> (a) The initial state of the mind, say at birth,
>
> (b) The education to which it has been subjected,
>
> (c) Other experience, not to be described as education, to which it has been subjected.

> Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain.

Tanken är med andra ord att först skapa en startpunkt, en barnmaskin, och sedan utbilda ”barnet” så att det motsvarar en vuxenmaskin. Turing delar upp problemet i två delar; att skapa startpunkten (barnet) och att skapa lärprocesser som utvecklar ”barnmaskinen” till ett vuxet stadium, samtidigt som han identifierar att det behövs någon form av urval för att skilja ut de maskiner som utvecklats framgångsrikt:

> We have thus divided our problem into two parts. The child programme and the education process. These two remain very closely connected. We cannot expect to find a good child machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications
>
> Structure of the child machine = hereditary material
>
> Changes of the child machine = mutation,
>
> Natural selection = judgment of the experimenter

Turing fortsätter med att reflektera över lärandeprocessen och bristen på kontroll över AI-programmets inre processer, något som inte minst kan vara viktigt ur ett beslutsperspektiv för att kunna förutse AI-programmets beteende i alla situationer.

> An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behavior. This should apply most strongly to the later education of a machine arising from a child machine of well-tried design (or programme). This is in clear contrast with normal procedure when using a machine to do computations: one's object is then to have a clear mental picture of the state of the machine at each moment in the computation. This object can only be achieved with a struggle. The view that "the machine can only do what we know how to order it to do,"' appears strange in face of this. Most of the programmes which we can put into the machine will result in its doing something that we cannot make sense of at all, or which we regard as completely random behaviour. Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops. Another important result of preparing our machine for its part in the imitation game by a process of teaching and learning is that "human fallibility" is likely to be omitted in a rather natural way, i.e., without special "coaching." […] Processes that are learnt do not produce a hundred per cent certainty of result; if they did they could not be unlearnt. (Turing, 1950, s.454)

Turings tankar om tillvägagångssätt vid skapande av artificiell intelligens; med en barnmaskin, eller ett embryo, som sedan utbildas för att dynamiskt kunna lära sig ett beteende och där förståelsen för de inre processerna ofta är begränsad har många paralleller till dagens användning av artificiell intelligens, som ofta delas upp i specialiserad/narrow eller generell artificiell intelligens. Inte minst identifierar Turing centrala frågeställningar och utmaningar gällande skapandet av lärande intelligenta maskiner. För det första ska någon form av entitet (ex program eller maskin) skapas med kapacitet att lära sig. Sedan ska entiteten utbildas, och effekten av utbildningen utvärderas.

I artikeln pekar även Turing ut en av de stora frågorna när det gäller självlärande system och artificiell intelligens, nämligen hur man kan skapa säkra system med tanke på att systemens inre processer ofta är mycket svåra att överblicka och förstå, en utmaning inte minst ur ett beslutsperspektiv – hur ska man kunna vara säker på att besluten är riktiga om man inte förstår hur de kommit till?

Turing avslutar ”Computing Machinery and Intelligence” (1950, s.455) med några funderingar om hur man skulle kunna gå vidare i utvecklingen av artificiell intelligens:

> We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.

Sammanfattningsvis ger med andra ord Turing förslag på två tänkbara utvecklingsvägar; antingen genom att specialisera sig på abstrakta uppgifter som exempelvis schackspel eller genom att ge AI-systemet sinnen och generell kapacitet för att kunna lära sig på ett liknande sätt som ett barn. Båda dessa strategier har provats med varierande framgång.

2.3 Specialiserad och generell artificiell intelligens

Sedan Turing publierade sina tankar om artificiell intelligens 1950 har det hänt mycket när det gäller både de tekniska lösningarna och den allmänna förståelsen av intelligens, inte minst i spåren av genombrott i förståelsen av mänsklig intelligens. Men utvecklingen kan ändå sägas i stora drag följa Turings tankar, exempelvis det gäller utforskandet av de två teknikstrategierna specialiserad respektive generell artificiell intelligens.

2.3.1 Specialiserad artificiell intelligens

Under flera år har utvecklingen framförallt varit fokuserad på att ta fram kretsar och program specifikt designade för att lösa ett visst problem, ofta kallat ”Narrow artificiell intelligence”, men i detta arbete fortsättningsvis kallad specialiserad artificiell intelligens. Exempelvis användes denna strategi i Deep Blue, som 1997 slog Kasparov i schack. Deep Blue är ett specifikt system för schackspelande, som bland annat var uppbyggt kring 4 000 olika spelöppningar samt utvärdering av varje tänkbart drag efter 8 000 parametrar med utgångspunkt i en databas innehållande 700 000 tidigare schackspel. Dessutom använde Deep Blue olika tumregler, heuristiker, för att så tidigt som möjligt kunna utesluta alternativ med låg potential (Nilsson 2010, s.592-594, Russell & Norvig 2010, s.185).

Ett annat exempel på specialiserad artificiell intelligens är IBMs datasystem Watson, som 2011 slog de mest framgångsrika mänskliga spelarna i det amerikanska frågespelet Jeopardy. Systemet är baserat på bland annat en stor databas över händelser och företeelser, men har också algoritmer för tolkning av kryptiskt formulerade frågor och värdera sannolikheten för olika svar. Utvecklarna bakom Watson skriver (Ferrucci *et al*., 2010):

> The system we have built and are continuing to develop, called DeepQA, is a massively parallel probabilistic evidence-based architecture. For the Jeopardy Challenge, we use more than 100 different techniques for analyzing natural language, identifying sources, finding and generating hypotheses, finding and scoring evidence, and merging and ranking hypotheses. What is far more important than any particular technique we use is how we combine them in DeepQA such that overlapping approaches can bring their strengths to bear and contribute to improvements in accuracy, confidence, or speed.

Sedan vinsten i Jeopardy har tekniken bakom systemet vidareutvecklats från att vara ett specialiserat system för frågesport till att bland annat arbeta med medicinska diagnoser, ge råd om matlagningsrecept samt styra autonoma fordon. Systemets kapacitet och inriktning kan därför idag sägas ha blivit betydligt mer generell än Deep Blue, även om inriktningen fortfarande på det stora hela tycks vara att snabbt bearbeta stora datamängder och sedan utifrån detta komma med ett svar eller en rekommendation. (Noyes, 2016)

2.3.2 Artificiell generell intelligens

Samtidigt som olika specialiserade tillämpningar av artificiell intelligens gjort stora framsteg under relativt lång tid så har utvecklingen av mer generella AI-system liknande biologiska hjärnor visat sig betydligt svårare än vad man tidigare trott, både när det gäller beräkningskraft och komplexitet.

Exempelvis var det först 2011 som IBM kunde meddela att de nu hade kapacitet nog att nå upp till en nivå motsvarande en katts hjärna, dvs 4,5 % av en normal mänsklig hjärna, även om kapacitet i sig knappast är det samma som att kunna göra en fullständig simulering (Fischetti, 2011). Sannolikt kommer det därför att dröja ett tag till innan vi har kapacitet nog att fullt ut kunna simulera en komplex biologisk hjärna.

Å andra sidan är det oklart hur mycket av hjärnans kapacitet som egentligen krävs för att kunna uppvisa mänskligt beteende. Detta illustreras exempelvis av ett fall med en man med skador på 90 % av hjärnvolymen som ändå kunde leva ett normalt liv med familj och arbete och, uppnå en IQ-poäng på 75 (Feuillet *et al*., 2007). Även om det i detta fall inte är klarlagt hur stor del av den viktiga hjärnbarken som de facto är förstörd, så är slutsaten att man rimligen ska vara lite försiktig med att uttala sig tvärsäkert om hur mycket kapacitet som egentligen krävs för intelligent beteende, något som även de flesta kattägare säkert instämmer i.

Men även om kapacitet och komplexitet fortfarande är betydande hinder för användning av mer generell artificiell intelligens så har det de senaste åren gjorts stora satsningar samtidigt som man har gjort betydande framsteg. OpenCog är ett projekt som syftar till att utveckla generell artificiell intelligens, baserat på Ben Goertzels tankar om intelligens. Enligt Goertzels modell kan intelligens i stort sett härledas till hierarkier av mönsterhantering och mönsterigenkänning, till system som dels organiserar inputs i mönster eller strukturer och dels som har funktioner som analyserar strukturerna för att identifiera mönster, och då även kontinuerligt utvecklas för att bli bättre och snabbare på att identifiera mönster (Cohen, 2016). Tanken är att man ska utveckla en grundstruktur som sedan kan utbildas genom mönster och mönsterigenkänning, helt i linje med Turings tankar. På sin hemsida skriver OpenCog (Pitt, 2010):

> In order to put this architecture to work, we have crafted a roadmap based on roughly mimicking the environment and development of young human children. A series of child-level learning tasks has been carefully laid out, which may be manifested via either virtual world agents or physical robots, and which lead from infant-level capabilities up to the grade school level. These tasks cover all the major cognitive capabilities displayed by young humans, and involve the integration of all major aspects of human intelligence, including perception, action, cognition, learning, memory, creativity, socialization, language, self-modeling, etc.

OpenCogs vision för projektet har till stor del handlat om att försöka skapa intelligenta robotar, som genom att interagera med människor sedan ska utveckla sin förmåga på ett sätt som påminner om barns mognad. Dessutom tänker man sig att robotarnas framsteg ska kunna göras tillgänglig för utvecklarna genom uppkoppling av robotarna till ett centralt nätverk. (Cohen, 2016)

Ett annat exempel på generell artificiell intelligens är det system som utvecklas av DeepMind, där man skapat ett system med en kapacitet som både är självlärande och generellt. Systemet bygger på en teknik som DeepMind kallar Deep Reinforcement Learning, som i sig är en kombination av de två teknikerna Deep learning och förstärkt inlärning (eng reinforcement learning). Deep learning är ett populärt begrepp för artificiell intelligens som baseras på flera neurala nätverk ordnade i hierarkiska nivåer, medan förstärkt inlärning är en inlärningsmetod där systemet kontinuerligt använder feedback om resultatet givet de olika handlingsalternativen och därigenom effektivt kan utvecklas och förbättras. Initialt användes gamla Atari-dataspel från 1980-talet som utvecklingsmiljö för den artificiella intelligensen, där systemet baserat på endast pixelinformation från skärmens ca 30 000 pixlar samt den aktuella

poängställningen själv lärt sig spela 50 olika spel som exempelvis Space invaders och Breakout på en övermänsklig nivå. (Hassabis, 2015, Mnih *et al*., 2015)

DeepMind har även utvecklat det system för generell artificiell intelligens som användes för att skapa AlphaGo, ett system för att spela brädspelet go. Systemet baseras på två neurala nätverk, ett policy-nätverk som utvecklades genom att systemet fick se över 100 000 olika amatörspel på internet och därigenom lärde sig att uppskatta sannolikheten för att en människa skulle göra ett visst motdrag givet den aktuella spelställningen. Detta användes sedan för att låta systemet fokusera på ett fåtal motdrag med störst sannolikhet, och därmed minska den nödvändiga beräkningskraften. Vidare skapade systemet ett värdenätverk som över tiden lärde sig förutspå utgången i spelet genom att låta policynätverket spela go mot sig själv 30 miljoner gånger. När AlphaGo 2015 besegrade europamästaren var systemet det första datorprogram som lyckades besegra professionell spelare utan handikapp. 2016 besegrade AlphaGo även Lee Sedol, som med 18 internationella titlar av många anses vara den bäste Go-spelaren det senaste årtiondet. (Hassabis, 2016a & 2016b)

Eftersom DeepMind-systemet till sin natur är generellt så kan det även lära sig många andra uppgifter. DeepMinds vision är att den artificiella intelligens som de utvecklar ska användas för att kunna lösa mänsklighetens utmaningar. Även om lösningen på många utmaningar fortfarande är utom räckhåll, så har DeepMinds system exempelvis använts för att optimera driften av datorhallar, något som gjort att Google kunnat minska kylningskostnaden i sina serverhallar med 40 % och den totala driftskostnaden med 15 %. (Clark, 2016, Hassabis, 2015, 2016a & 2016b)

2.3.3 Superintelligens

Eftersom utvecklingen av både teknik och algoritmer kontinuerligt gör framsteg, så är det rimligt att anta att vi en dag har tillgång till en generell artificiell intelligens som är kraftfullare än den mänskliga hjärnan, något som brukar kallas ”Superintelligence” eller superintelligens på svenska (Bostrom, 2014). Detta är något som man inom framförallt den filosofiska diskussionen om artificiell intelligens på senare år har intresserat sig allt mer för. En farhåga som ofta nämns är att tekniken då kan användas för att förbättra artificiell intelligens ytterligare och på så sätt kraftigt öka utvecklingstakten för artificiell intelligens. Denna aspekt presenterades först av Vernor Vinge, som döpte den till ”teknisk singularitet” och beräknas av Googles utvecklingschef Raymond Kurzweil inträffa år 2045 (Vinge, 1993, Grossman, 2011). Å andra sidan har en studie av träffsäkerheten i prognoserna kring artificiell intelligens visat att denna typ av förutsägelser generellt har anmärkningsvärt låg träffsäkerhet (Armstrong & Sotala, 2015). I vilket fall så ligger utvecklingen av superintelligens sannolikt många år bort, varför frågeställningar som rör detta område behandlas mycket översiktligt i denna uppsats.

Sammantaget kan man säga att den vision som Turing (1950) beskrev gällande utvecklingen av artificiell intelligens för över 60 år sedan och som kan betraktas som födelsen av artificiell intelligens fortfarande är aktuell och relevant. Utvecklingen kan fortfarande i grova drag beskrivas som uppdelad i två olika huvudspår; dels de som utvecklar kretsar och program för att lösa ett specifikt beslutsproblem så bra som möjligt, och dels de som försöker utveckla ett system med en generell intelligens som kan appliceras på många olika typer av beslutssituationer. Eftersom teknikutvecklingen kontinuerligt gör framsteg menar jag att det är rimligt att anta att vi en dag kommer att ha tillgång till artificiell intelligens med en kapacitet som överstiger den mänskliga hjärnan, något som ofta benämns ”superintelligens”.

2.4 Fem olika funktionsmodeller för artificiell intelligens

Sedan Turing skrev sin artikel så har många olika tekniker för artificiell intelligens provats, både för specialiserade intelligenta lösningar och generell artificiell intelligens. Inspiration till olika tekniker kan spåras till många olika områden, exempelvis neurovetenskap, datavetenskap, statistik, logik och evolutionslära. Med utgångspunkt i Pedro Domingos bok ”The Master Algorithm” (2015) kommer jag här att övergripande presentera fem olika familjer av tekniker som använts för att skapa artificiell intelligens; symbolistiska (symbolists), nätverkande (connectionists), evolutionära (evolutionaries), Bayesiska (Bayesians), och analogistiska (analogizers).

2.4.1 Symbolistiska metoder

Enligt den symbolistiska ansatsen kan intelligens skapas genom bearbetning av symboler, exempelvis genom matematiska ekvationer som kan bearbetas och lösas genom att ersätta ett uttryck med ett annat (jfr algebra).

> For symbolists, all intelligence can be reduced to manipulating symbols, in the same way that a mathematician solves equations by replacing expressions by other expressions. Symbolists understand that you can’t learn from scratch: you need some initial knowledge to go with the data. They’ve figured out how to incorporate preexisting knowledge into learning, and how to combine different pieces of knowledge on the fly in order to solve new problems. Their master algorithm is inverse deduction, which figures out what knowledge is missing in order to make a deduction go through, and then makes it as general as possible. (Domingos, 2015, s.52)

Symbolistiska metoder för artificiell intelligens har en lång tradition inom området, exempelvis beskrev John Haugeland (1985, s.112) tekniken med akronymen GOFAI – ”Good Old-Fashioned Artificial Intelligence” i sin bok Artificial Intelligence: The Very Idea från 1985. Datorn kan ses som en universell symbolmaskin (Janlert, 2015, s.67) och ansatsen att skapa intelligent beteende genom att bearbeta symboler är därför en logisk och närliggande metod.

En relativt vanlig symbolistisk metod inom artificiell intelligens är att använda regressionsanalys för att omforma ett antal mätpunkter till ett linjärt samband och sedan använda detta samband som en beslutsfunktion. I takt med att nya mätpunkter tillkommer kan sedan systemet självt anpassa sitt beteende till en föränderlig omvärld.

En styrka med symbolistiska metoder är deras enkelhet och transparens, då metoderna baseras på en linjär målekvation som kan analyseras och verifieras. Detta är samtidigt också en svaghet, eftersom symbolistiska ansatser därmed inte kan hantera ickelinjära samband eller undantag.

2.4.2 Nätverkande metoder

En annan ansats är att organisera kunskap och analys i olika former av nätverk, och skapa intelligent beteende genom att förstärka eller försvaga kopplingarna mellan nätverkens noder. Ett exempel är sk artificiella neurala nätverk, som inspirerats av neurovetenskapens tankar om neuronens och hjärnans funktion, inte minst av Hebbs regel (som tidigare nämnts):

> Neurons that fire together, wire together. Neurons that fire out of sync, lose their link (Munz *et al*., 2014).

Ett artificiellt neuralt nätverk består av ett stort antal enklare enheter (ex algoritmer) som är förbundna med länkar, där länkarna förstärks eller försvagas analogt med Hebbs regel (Janlert, 2015, s.153). Detta skapar ett system som beter sig på ett sätt som kan sägas vara plastiskt eller responsitivt i förhållande till omvärlden.

> For connectionists, learning is what the brain does, and so what we need to do is reverse engineer it. The brain learns by adjusting the strengths of connections

> between neurons, and the crucial problem is figuring out which connections are to blame for which errors and changing them accordingly. The connectionists' master algorithm is backpropagation, which compares a system's output with the desired one and then successively changes the connections in layer after layer of neurons so as to bring the output closer to what it should be. (Domingos, 2015, s.52)

Precis som symbolistiska metoder så har även neurala nätverk en lång tradition inom artificiell intelligens och på senare tid har tekniken fått ett kraftigt genomslag, inte minst genom att man kombinerar flera neurala nätverk i en hierarkisk struktur – något som ofta kallas ”Deep Learning”. Exempelvis är denna teknik en central del i DeepMinds ansats till generell artificiell intelligens och har även visat sig vara effektiv vid exempelvis bildtolkning, där tekniken framgångsrikt använts till att både identifiera vad den föreställer och var bilden är tagen. (Hassabis, 2016b, Weyand *et al*., 2016)

En fördel med nätverksansatser framför exempelvis symbolistiska är att tekniken inte förutsätter linjära samband, utan därför har lättare att hantera undantag och ickelinjära målfunktioner och samband. Samtidigt medför detta att de målfunktioner som är resultatet av tillämpningar med artificiell intelligens baserad på en nätverkande ansats tenderar att vara svåra att överblicka och analysera (Janlert, 2015, s.154).

2.4.3 Evolutionära metoder

Med utgångspunkt i de tankar om det naturliga urvalet och ”survival of the fittest” som Charles Darwin presenterade i ”On the origin of species” (1861) har man även skapat evolutionära metoder för bland annat programvaruutveckling och tekniker för utveckling och förbättring av beslutsalgoritmer inom artificiell intelligens. Först låter man de grundläggande algoritmerna genomgå en slumpmässig påverkan och därefter utvärderas de olika varianterna och de mest framgångsrika kombineras till nya ”barn” som ärver egenskaper från sina ”föräldrar” enligt evolutionära principer, varefter processen oftast återupprepas i flera led.

> Evolutionaries believe that the mother of all learning is natural selection. If it made us, it can make anything, and all we need to do is simulate it on the computer. The key problem that evolutionaries solve is learning structure: not just adjusting parameters, like backpropagation does, but creating the brain that those adjustments can then fine-tune. The evolutionaries' master algorithm is genetic programming, which mates and evolves computer programs in the same way that nature mates and evolves organisms. (Domingos, 2015, s.52)

Även Nilsson (2010, s.43) tar upp evolutionära mekanismer som en tänkbar metod för att utveckla artificiell intelligens:

> That living things evolve gives us two more clues about how to build intelligent artifacts. First, and most ambitiously, the processes of evolution itself – namely, random generation and selective survival – might be simulated on computers to produce the machines we dream about. Second, those paths that evolution followed in producing increasingly intelligent animals can be used as a guide for creating increasingly intelligent artifacts. Start by simulating animals with simple tropisms and proceed along these paths to simulating more complex ones. Both of these strategies have been followed with zest by AI researchers.

Evolutionära processer har bland annat visat sig vara relativt framgångsrika för att i kombination med andra tekniker kunna generera nya och kraftfulla beslutsalgoritmer. Tekniken har exempelvis visat sig kunna kombineras med både symbolistiska och nätverkande ansatser, för att på det sättet generera allt kraftfullare beslutsalgoritmer för varje ny generation. Ett exempel är de simuleringar av tvåbenta datormodeller som Torsten Reil demonstrerade i ett TED-talk 2003, där han visade hur man kan använda sig av en evolutionär metod för att utveckla ett neuralt nätverk för kontroll av rörelse och balans (Riel, 2003). Samtidigt leder evolutionära principer ofta till mycket komplexa och svårbegripliga beslutsalgoritmer eftersom komplexiteten ofta ökar med

varje ”generation”. Detta gäller givetvis särskilt om metoden kombineras med en metod som i sig själv är svår att överblicka, som exempelvis ett neuralt nätverk.

2.4.4 Bayesiska nätverk

En annan teknik som ofta används inom artificiell intelligens är Bayesiska nätverk, som baseras på Bayes teorem om betingade sannolikheter, som är central inom sannolikhetsläran. Bayesiska nätverk kan därför sägas vara är särskilt lämpade vid tillämpningar som kräver hantering av osäkerhet.

> Bayesians are concerned above all with uncertainty. All learned knowledge is uncertain, and learning itself is a form of uncertain inference. The problem then becomes how to deal with noisy, incomplete, and even contradictory information without falling apart. The solution is probabilistic inference, and the master algorithm is Bayes' theorem and its derivates. Bayes' theorem tells us how to incorporate new evidence into our beliefs, and probabilistic inference algorithms do that as efficiently as possible (Domingos, 2015, s.52).

Även Nilsson (2010) tar upp Bayesiska nätverk och teknikens betydelse vid situationer som kräver hantering av osäkerhet, något som ofta är centralt vid beslutsfattande.

> Because nearly all reasoning and decision making take place in the presence of uncertainty, dealing with uncertainty plays an important role in the automation of intelligence. Attempts to quantify uncertainty and “the laws of chance” gave rise to statistics and probability theory. What would turn out to be one of the most important results in probability theory, at least for artificial intelligence, is Bayes's rule (Nilsson, 2010, s.52).

Användningsområden finns inom medicin, bildbehandling och beslutsstödsystem, bland annat för skräpposthantering eller mönsterigenkänning (Yusof & Mokhtar, 2016). Fördelar ligger i att en stor mängd data kan behandlas snabbt och kostnadseffektivt. En nackdel är att svaret blir approximativt. Ett exempel på ett ai-system som delvis baseras på denna metod är IBMs Watsonsystem, där vägningen av sannolikheten för olika svarsalternativ är central vid systemets analys.

2.4.5 Analogistiska metoder

Den femte ansatsen för artificiell intelligens som nämns av Domingos är analogistiska metoder, som framförallt fokuserar på att analysera likhet:

> For analogizers, the key to learning is recognizing similarities between situations and thereby inferring other similarities. If two patients have similar symptoms, perhaps they have the same disease. The key problem is judging how similar two things are. The analogizers' master algorithm is the support vector machine, which figures out which experiences to remember and how to combine them to make new predictions (Domingos, 2015, s.53).

Analogistiska metoder har som jag ser det påtagliga paralleller med Ben Goertzels tankar om intelligens. Som presenterades under avsnittet om generell artificiell intelligens så härleder Goertzel intelligens till mönsterigenkänning och hierarkier av mönster vilket som jag ser det har tydliga paralleller till den analogistiska ansatsen.

Kognitionsforskaren och författaren Douglas Hofstadter är en annan förespråkare av detta synsätt. I sin bok Gödel, Escher, Bach: An Eternal Golden Braid skriver exempelvis Hofstadter (1979, s.709) att han betraktar mönsterigenkänning som kärnan i såväl artificiell intelligens som förståelsen av den mänskliga kognitionen.

2.4.6 Kombinationer av olika metoder

De olika metoderna kan både användas enskilt eller i olika kombinationer med varandra. Evolutionära metoder kan exempelvis användas för att få fram bättre symbolistiska målekvationer eller neurala nätverk.

Ett annat exempel är sk Deep Learning, som kan ses som ett samlingsnamn för metoder som består av flera olika nivåer av ickelinjär informationsbearbetning samt metoder för övervakad (jfr eng supervised) eller oövervakad inlärning på varje nivå. Denna typ av ansatser för artificiell intelligens har på senare år gjort betydande framsteg, drivet av bland annat allt kraftfullare datorchip (inte minst genom användningen av olika typer av beräkningskretsar exempelvis baserade på grafikprocessorer) och användandet av allt större datamängder vid maskininlärningen. (Deng & Yu, 2014, s.201)

Exempelvis består den generella artificiella intelligens som används av DeepMind i AlphaGo av två neurala nätverk, där systemet själv lärt sig att analysera värdet av olika drag genom att studera hundratusentals partier och jämföra positionerna med det slutliga utfallet. Denna typ av ansats tycks ofta ha ett betydande prestandaövertag jämfört med exempelvis symbolistiska metoder samtidigt som lösningarna i sig därmed medför ökad komplexitet och därmed är svårare att utvärdera. (Hassabis, 2016a, 32:40)

2.5 Tre generella metoder för maskininlärning

Utöver en metod för den artificiella intelligensens logiska struktur, något som diskuterades ovan, behövs även en metod för hur systemet ska lära sig av indata för att systemet ska bete sig intelligent. För detta finns många olika metoder som kan delas i tre huvudkategorier; övervakade (eng supervised), oövervakade (eng unsupervised) och förstärkt (reinforcement) inlärning (Russell & Norvig, 2010, s.694-695).

2.5.1 Övervakade metoder

Vid övervakade metoder så får AI-systemet ledtrådar till önskvärda beslut genom att varje post i tränings- och testdata data får en värdering i form av en poäng eller ett omdöme. Systemets uppgift är sedan att försöka identifiera en modell, exempelvis en algoritm eller ett neuralt nätverk, som beskriver kopplingen mellan indata och värderingen. Sedan testas modellen mot en uppsättning testdata som precis som träningsdata är värderad för att man ska kunna utvärdera modellens träffsäkerhet.

2.5.2 Oövervakade metoder

Vid oövervakade metoder får systemet analysera datamaterialet och själv försöka identifiera en så effektiv gruppering av datamaterialet som möjligt, vilket exempelvis sedan kan användas för att analysera likhet. Inlärningsprincipen kan därför vara användbar vid analogistiska metoder och kan t.ex. användas för klassificeringar eller diagnoser, något som kan vara till stor hjälp i system som generar beslutsunderlag medan metoden är svårare att använda för ett aktivt beslut eftersom systemet i sig självt inte har någon kunskap om vad som är ett önskvärt utfall.

2.5.3 Förstärkt inlärning

Vid förstärkt inlärning skapas först en beslutsfunktion, exempelvis en algoritm eller ett neuralt nätverk, baserad på indata. Därefter används beslutsfunktionen på nya data och systemet får kontinuerligt positiv och negativ feedback beroende på utfallet. Utefter denna feedback uppdateras beslutsfunktionen, så att systemet hela tiden försöker förbättra sin träffsäkerhet baserat på prestationen.

2.6 Funktion och inlärning ur ett beslutsperspektiv

Som framgår av presentationerna av de olika logiska metoderna, så finns det många olika ansatser och tekniker som var för sig eller i kombination kan användas för att skapa artificiell intelligens. Alla tekniker för funktion och inlärning har sina styrkor och svagheter, och passar därför olika bra vid olika typer av tillämpningar.

2.6.1 Funktionsmodellerna ur ett beslutsperspektiv

Som jag ser det kan de fem olika funktionsmodellerna grupperas i två huvudkategorier baserat på både deras ursprung och generella karaktär; matematiska och organiska. Bland de matematiskt inspirerade teknikerna hittar vi symbolistiska metoder samt Bayesiska nätverk, medan de nätverkande och evolutionära teknikerna kan ses som mer organiska eller biologiskt inspirerade. I mina ögon framstår den analogistiska metoderna som något av en hybrid, med paralleller till både matematik och biologi.

Gemensamt för de matematiskt inspirerade metoderna är inte bara den bakomliggande inspirationskällan, utan de delar även några generella drag – som att de till sin uppbyggnad är mer generella, regelstyrda och därför resulterar i beslutsalgoritmer som är enklare och därför ofta framstår som relativt förutsägbara och mer transparenta. De symbolistiska ansatserna ger i regel upphov till en målekvation, som kan testas och verifieras, medan Bayesiska nätverk exempelvis kan åskådliggöras i ett eller flera träddiagram. Samtidigt så gör de matematiskt inspirerade metodernas regelbaserade struktur att de har svårt att hantera undantag eller ickelinjära samband.

När det gäller de biologiskt inspirerade metoderna, så tycks dessa generellt vara mer kraftfulla än de matematiska samtidigt som de genererar beslutsfunktioner som i regel är betydligt mer komplexa och därmed svåra att tyda. Nätverksansatser som exempelvis neurala nätverk har fått ett betydande genomslag på senare år och har sin styrka i bland annat de situationer som är svåra eller omöjliga att precisera i en linjär beslutsfunktion, exempelvis vid bildtolkning. Evolutionära ansatser är ett kraftfullt verktyg för att ”odla” fram nya symbolistiska beslutsekvationer eller beslutsnätverk i enlighet med de evolutionära principerna arv, miljö och ”survival of the fittest”.

Analogistiska metoder, som utgår från likhet eller mönsterigenkänning, är en intressant ansats som har påtagliga likheter med den biologiska kognitionen samtidigt som implementationen genom exempelvis vektoranalys kan göras matematiskt effektiv. Samtidigt så är likhet med biologiska funktionssätt i sig ingen garanti för att det är en effektiv ansats för artificiell intelligens. Analogt med när människan lärde sig flyga genom att överge de fladdrande biologiska vingarna och i stället att separera fart (motorer) och bärkraft (fixerade vingar) kan man tänka sig att vägen till funktionell artificiell intelligens går genom att separera, renodla och kombinera olika funktioner istället för att försöka replikera de biologiska lösningarna.

Sammantaget kan man konstatera att det finns många olika metoder för artificiell intelligens, som skiljer sig såväl när det gäller effektivitet som transparens. När man väljer vilken eller vilka ansatser till artificiell intelligens som är lämpliga vid en specifik beslutstillämpning är det därför viktigt att noga överväga faktorer som effektivitet, transparens, linjäritet och förutsägbarhet, och anpassa valet av metod till den aktuella situationen. I många tillämpningar är sannolikt beslutskraften hos symbolistiska metoder fullt tillräcklig, exempelvis i specialiserade system som syftar till beslutsavlasting vid rutinmässiga beslut. I mer komplexa beslutssituationer eller för generell artificiell intelligens är däremot mer kraftfulla och flexibla metoder sannolikt att föredra, exempelvis kombinationer av ickelinjära olika metoder, sk Deep learning.

2.6.2 Inlärning ur ett beslutsperspektiv

Ur ett beslutsperspektiv kan man se att de olika metoderna för inlärning fyller olika funktioner och leder till olika beteenden. Vid övervakade metoder har systemet tydligt fått en värdering för varje datapost, vilket gör att metoden passar bra för att användas för olika typer av beslutsfunktioner. Samtidigt så kan denna värdering i sig vara kontroversiell och leda till frågor om de bakomliggande värderingarna, exempelvis om dessa kan sägas vara etiska eller rättvisa.

Motsatsen är oövervakade metoder, som fungerar som en klassificering eller gruppering och kan därför vara ett utmärkt analysverktyg för att gruppera ett obearbetat datamaterial för vidare analys. Resultatet i sig självt saknar värdering, vilket å ena sidan gör att man kan säga att analysen är neutral i förhållande till värdepåverkan, men samtidigt är systemet i sig inte särskilt användbart för aktivt beslutsfattande där systemet på något sätt behöver ”förstå” vad som är ett önskvärt resultat.

Förstärkt inlärning utgår från en beslutsfunktion och använder feedback från utfallet för att låta funktionen förbättras. Denna metod har mycket gemensamt med exempelvis mänsklig inlärning, eftersom vi oftast får feedback på resultatet snarare är det precisa utförandet.

Ur ett beslutsperspektiv kan därför metoder besläktade med förstärkt inlärning vara passande när man vill nå ett visst resultat och inte har någon värdering i hur resultatet uppnås. Övervakade metoder ger större kontroll över själva beteendet, vilket både kan vara en fördel och en nackdel. En fördel eftersom det öppnar för mer kontroll över systemet, så att exempelvis beslut baserade på värderingar som vi uppfattar som oetiska (exempelvis rasistiska eller sexistiska) kan undvikas. Men samtidigt en nackdel eftersom den förutsätter att varje datapost ska fått ett omdöme som i sig kan vara snedvridet eller oetiskt. Dessutom riskerar system baserat på övervakad inlärning att bli mindre effektivt jämfört med ett system baserat på förstärkt inlärning, eftersom den senare metoden endast försöker identifiera den effektivaste lösningen och inte tar hänsyn till hur denna uppnås.

3 AI som beslutsmetod idag

Utvecklingen inom artificiell intelligens har på senare tid tagit stora steg. För att ge en indikation av utvecklingen kommer jag här att presentera några generella utvecklingstrender baserade på patentdatabasen Espacenet (EPO, 2017), beskriva utvecklingen inom fyra specifika tillämpningsområden samt kortfattat beröra några av tekniskens begränsningar.

3.1 Generella utvecklingstrender

På senare år har den snabba utvecklingen inom artificiell intelligens överraskat många, även bland annat de som står utvecklingen nära. Exempelvis berättade Googles medgrundare Sergey Brin på World Economic Forum om sin tidigare inställning till artificiell intelligens (WEF, 2017):

> “I didn’t pay attention to it at all, to be perfectly honest. Having been trained as a computer scientist in the 90s, everybody knew that AI didn’t work. People tried it, they tried neural nets and none of it worked.”

Ett sätt att åskådliggöra den generella utvecklingen inom artificiell intelligens är använda den globala patentdatabasen Espacenet (EPO, 2017). Av denna framgår exempelvis att termen ”machine learning” används för första gången i en patenttitel 1962 och fram till 1980 förekommer termen 5 gånger inom IPC-kategorin G06 (Computing, Calculation, Counting). Därefter följer en utveckling som i det närmaste kan beskrivas som exponentiell:

Tabell 1. Antal patent inom IPC-kategorin G06 med termen ”Machine learning” i titeln, baserat på en sökning i Espacenet 20170220.

-1979	5
1980-1989	37
1990-1999	140
2000-2009	278
2010-	1051
Totalt: 1534, varav USA 698, Kina 385 och Sverige 5	

Om man istället söker på termen neurala nätverk (neural networks) får man 859 träffar där termen använts i titeln, varav den första är från 1985. Från 1990 till 2009 tycks utvecklingstakten varit relativt jämn, för att sedan accelerera, med exempelvis över 100 publicerade patent under 2016:

Tabell 2. Antal patent inom IPC-kategorin G06 med termen ”Neural networks” i titeln, baserat på en sökning i Espacenet 20170220.

-1979	0
1980-1989	12
1990-1999	259
2000-2009	249
2010-	356
Totalt: 859, varav USA 503, Kina 27 och Sverige 0	

Vid motsvarande sökning inom termen ”Deep learning” påträffas termen i en patenttitel första gången 2009, och finns totalt i titeln på 221 patent:

Tabell 3. Antal patent inom IPC-kategorin G06 med termen ”Deep learning” i titeln, baserat på en sökning i Espacenet 20170220.

2009	1
2010	1
2011	0
2012	2
2013	3
2014	18
2015	73
2016	105
2017	11
Totalt: 221, varav USA 31, Kina 154 och Sverige 0	

Som datakälla har en patentdatabas som Espacenet flera svagheter, exempelvis eftersom olika länder har olika praxis för hur patenten ska registreras, samtidigt som summorna mellan de olika årtalen inte matchar årtalen (totalen som redovisas av databasen är större än summan av de olika delarna). En annan aspekt att ta hänsyn till är att ett patent i vissa fall är registrerat flera gånger, beroende på vilken region som det gäller samt att det i systemet finns en betydande eftersläpning mellan det att ansökan registreras, till dess att det beviljas och därefter publiceras. Givet dessa svagheter bör de redovisade sifforna tolkas som ett stickprov som ger en indikation på den generella utvecklingen.

Men givet denna invändning finns ändå göra flera intressanta observationer. En är att ökningen av patent relaterade till artificiell intelligens tycks ha en i det närmaste exponentiell ökning sedan de första patenten på 60-talet. Inte minst är ökningen sedan 2010 anmärkningsvärt kraftig.

En annan observation är att det tycks finnas stora regionala skillnader mellan olika länder/regioner i användningen av olika termer för besläktade tekniker, exempelvis neural networks och Deep learning. Detta kan vara en konsekvens av skillnader i språkbruk, men sannolikt tyder det på att regionerna tycks ha lite olika inriktning i sin utveckling. Exempelvis tycks Kina framförallt inrikta mycket av utvecklingen inom artificiell intelligens på komplexa hierarkiska kombinationer av ickelinjära metoder, sk Deep learning, och står för 70 % av de beviljade patenten med koppling till detta begrepp. Annars tycks USA ha en särställning inom kopplade relaterade till artificiell intelligens. Bland patenten kopplade till machine learning står landet för 46 % av ansökningarna, medan motsvarande siffra är 59 % inom neural networks.

Vidare kan USAs och Kinas dominans i patentstatistiken tyda på att exempelvis Europa riskerar att hamna på efterkälken när det gäller utvecklingen av artificiell intelligens. Detta intryck förstärks även av att det vid sökningarna endast påträffades ett fåtal patent med svenska upphovsmän registrerat i Espacenet, exempelvis fem stycken relaterade till machine learning (varav fyra tycks bygga på samma grundansökan). Om dessa siffor är en korrekt avspegling av utvecklingen inom området så kan Europa och Sverige stå inför betydande utmaningar i den konkurrensutsatta globala ekonomin.

3.2 Några tillämpningar av AI som beslutsmetod idag

Tillämpningarna av artificiell intelligens kan sägas kontinuerligt bli både bredare och djupare, bredare eftersom tekniken används på allt fler områden och djupare eftersom tillämpningarna hela tiden blir allt mer avancerade. Jag kommer här att kort presentera några av dagens tillämpningar av artificiell intelligens inom fyra specifika områden; i trafiken, inom finanssektorn, inom medicinen och inom militären, som alla är exempel på områden där det pågår en intensiv utveckling av olika tillämpningar med artificiell intelligens.

3.2.1 Autonoma fordon

I trafiken möter vi många situationer som är svåra för vår hjärna att hantera på ett bra sätt. Exempelvis behöver vi kontinuerligt ha hög uppmärksamhet och mental närvaro med ett snabbt beslutsfattande baserat på en stor mängd av olika sorters inkommande information för att undvika olyckor, något som är svårt för en hjärna som har relativt lätt för att bli överbelastad, distraherad eller trött. Som tidigare nämnts fann en studie vid Uppsala universitet att ”den mänskliga faktorn” var inblandad i 90-95 % av alla trafikolyckor (Forward, 2008). Med tanke på att det enligt transportstyrelsens olycksstatistik dör mellan 250 och 400 personer i den svenska trafiken år, och mellan 17 000 och 26 000 skadas, så är det lätt att se den potentiella nyttan av säkrare självkörande bilar (Transportstyrelsen, 2016).

För att avlasta fören har det därför på senare år blivit allt vanligare med olika typer av stöd för föraren, allt ifrån GPS-system som hjälper föraren att hitta vägen till olika former av aktiva säkerhetssystem som exempelvis adaptiva farthållare som anpassar farten till andra trafikanter, anti-kollisionssystem med nödbromsning och system som kontrollerar uppmärksamheten och tröttheten hos föraren (Roos, 2010).

Men i takt med de tekniska framstegen inom datorer, sensorer och robotstyrning så har även teknik för autonoma fordon utvecklats mycket snabbt och det pågår ett flertal olika projekt med autonoma fordon, både bland traditionella bilföretag som Audi och Volvo, men även Google och Uber. Dessutom är bilar som elbilstillverkaren Tesla levererar redan försedda med all hårdvara som krävs för att fordonet ska kunna köra självt, men funktionsmässigt kan systemet idag snarast liknas vid en förhållandevis

avancerad autopilot som exempelvis övervakar avstånd till andra bilar och kan byta motorvägsfil själv. Även lastbilstilverkare som Scania ser stora fördelar med autonoma lastbilar och investerar mycket i tekniken. (Google, 2016, Törnros, 2015, Tesla, 2016, Volkswagen Group Sverige AB, 2016, Wemmenhag, 2016)

3.2.2 Finansiella handelsrobotar och robotrådgivare

På de finansiella marknaderna hanteras enorma belopp otroligt snabbt, samtidigt som den mänskliga hjärnan är relativt långsam och de medvetna tankeprocesserna är relativt krävande (Kahneman, 2013), något som bland annat medfört omfattande användning av datasystem som styrs av olika beslutsalgoritmer inom aktie- och valutahandel. Dessutom genererar den ekonomiska världen hela tiden enorma mängder statistik och annan ekonomisk data, något som kan vara svårt att överblicka för en människa, medan datorer gör det betydligt effektivare. Detta märks bland annat i handelsstatistiken på världen börser, exempelvis stod algoritmhandeln för 45 % och högfrekvenshandeln för 13 % av all handel på stockholmsbörsen 2011 enligt Nasdaq OMX Stockholm (Sveriges Riksbank, 2015).

Algoritmstyrd aktiehandel har funnits på världens börser sedan åtminstone slutet av 90-talet, då börserna gick över till digitaliserade handelsplattformar som kan nås via internet. Den tekniska utvecklingen möjliggjorde olika former av algoritmbaserade handelssystem, där automatiserade datasystem fattar handelsbeslut baserat på handelsmönster och aktuell orderbok. Ett sådant exempel är den sk högfrekvenshandeln, som bygger på handelsfördelar genom extremt snabba uppkopplingar till börserna, för att på det sättet förekomma handelsrörelser och eftersom en högfrekvenshandlarare kan beordra mer än 100 000 handelsordrar per sekund så är handeln till sin natur alldeles för snabb för den mänskliga hjärnan utan styrs av förbestämda, oftast relativt statiska, beslutsalgoritmer. (Buchanan, 2015)

Dessa programmerade handelssystem för värdepappershandel innefattar system med olika grader av autonomi i de datoriserade systemen. I algoritmiskt programmerad handel innehåller programmen avancerade matematiska formler. Dessa formler kan bland annat analysera hur handeln påverkar priserna på aktierna, ofta för att kunna genomföra stora transaktioner utan att orsaka prisförändringar som leder till paniska köp-eller säljrusher på börsen. Ofta är programvaran adaptiv på ett mer eller mindre avancerat sätt och utvecklar sitt beteende genom den kontinuerligt utvärderar sina egna misstag och möjligheter. (Salmon & Stokes, 2010, Söderström 2011)

På senare år har det dock blivit vanligare med handelssystem som uttalat är baserade på artificiell intelligens, exempelvis Hong Kong-baserade Aditya, som startats av Ben Goertzel, som nämnts tidigare i arbetet. Ett annat exempel är Sentient, som grundats av bland andra Babak Hodjat, som även utvecklat Apples digitala assistent Siri (Kiat, 2016). Senitent beskriver själva sin tillämpning av artificiell intelligens som en kombination av evolutionär artificiell intelligens och Deep Learning (dvs en hierarki av ickelinjär informationsbearbetning, exempelvis genom neurala nätverk), tekniker som de använder för att bearbeta stora datamängder och identifiera intressanta handelsstrategier, en tillämpning som ofta kallas ”Big Data” (Senitent, 2016). I takt med att tekniken blivit allt billigare så har denna typ av tillämpningar blivit allt vanligare även bland mindre investerare, och på Internet är det idag relativt enkelt att hitta beskrivningar på handelsapplikationer baserade på artificiell intelligens som exempelvis neurala nätverk och källkod för att skapa sin egen handelsrobot, exempelvis på sidor som www.investopedia.com (Vonko, 2016). Även för privatpersoner kommer det olika former av rådgivningstjänster som baseras på artificiell intelligens, exempelvis tjänster som Wealthfront, Betterment, Stellum, Lifeplan och Primepilot (Ström, 2016). Även om dessa tjänster ännu så länge framförallt tycks vara inriktade på rådgivningen så är rena placeringstjänster baserade på artificiell intelligens sannolikt mycket närliggande.

3.2.3 Medicinska tillämpningar

Även inom medicinen finns det en betydande potential för användning av artificiell intelligens, exempelvis inom diagnostik. En studie från Socialstyrelsen indikerade exempelvis att misstag och felbedömningar årligen leder till över 100 000 vårdskador, som i sig bidrar till över 3 000 dödsfall och leder till 630 000 extra vårddygn (Socialstyrelsen, 2008). En av vårdens utmaningar är att ställa rätt diagnos och ge rekommendation om bästa möjliga behandling baserat på all den forskning som hela tiden publiceras vid världens universitet. Inom området publiceras dagligen ca 8000 nya vetenskapliga artiklar, något som gör det mycket svårt för läkare att hålla sig uppdaterad på den senaste forskningen. (Rose, 2016)

Vid University of North Carolina at Chapel Hill pågår därför försök att använda IBMs Watson-system för artificiell intelligens inom cancervård för att både förbättra diagnostiken och ha bättre överblick över aktuell forskning. Watson-systemet utvecklades från början för att nå framgångar i spelet Jeopardy genom att bland annat snabbt kunna bearbeta stora mängder text och baserat på detta snabbt komma upp med ett relevant svar, egenskaper som gör att systemet effektivt kan gå igenom databaser med forskningsrapporter. Dessutom har systemet kompletterats med bildanalys, så att det kan ge förslag på diagnoser och behandlingar baserat på röntgenbilder tillsammans med olika provsvar. Försöket med artificiell intelligens som diagnosverktyg har hittills varit mycket framgångsrikt. Vid en utvärdering av systemet baserad på 1000 patienter fann man att systemet i 99 % av fallen gjorde samma bedömning som de behandlande läkarna, men att systemet i 30 % också kom med rekommendationer på behandlingar som de behandlande läkarna missat. (Steadman, 2013, Rose, 2016)

Andra exempel på användning av artificiell intelligens inom vården är företagen Verscend och HealthTap. Verscend hjälper bland annat amerikanska försäkringsbolag att utvärdera vårdkostnader och vårdeffektivitet baserat på dataanalys med hjälp av artificiell intelligens (Verscend, 2016). HealthTap använder artificiell intelligens för effektivare medicinsk rådgivning baserat på råden från över 100 000 läkare, något som både medför att fler även i fattigare delar av världen får tillgång till kvalificerad rådgivning och en potentiell rationalisering av vägledningsfunktionen inom vårdsektorn, en funktion som i Sverige idag löses genom allmänläkare och vårdcentraler (HealthTap, 2014).

3.2.4 Autonoma vapensystem

Inom militären har man i flera år använt artificiell intelligens för att lösa administrativa uppgifter som exempelvis planering av logistikflöden. I samband med det första Irakkriget 1991 exempelvis användes ett logistiksystem baserat på artificiell intelligens som på några timmar genererade fraktplaner för bland annat 50 000 fordon, fraktgods och personal, något som med traditionella metoder hade tagit veckor (Russell & Norvig, 2010, s.29)

På senare år har det även blivit allt vanligare med olika former av militära robotar för svåra eller farliga uppdrag, exempelvis iRobots Warrior och Packbot samt General atomics Reaper- och Predator-drönare. Även om dessa idag i huvudsak är fjärrstyrda så finns omfattande planer på att i nästa generation förse dem med funktioner baserade på artificiell intelligens. Vid tester av sådana system tycks dessa redan idag vara så välutvecklade att de är överlägsna sina mänskliga motsvarigheter i stridssituationer. (Chamary, 2016)

Men redan idag finns det olika autonoma vapensystem i bruk; exempelvis Sydkoreanska Sentry och Super aEgis II, samt Israeliska Harop. Sentry är ett robotsystem utvecklat av Samsung beväpnat med maskingevär och granatkastare som använder rörelse- och värmesensorer för att upptäcka fientlig rörelse i den demilitariserade zonen mellan Nord- och Sydkorea. När Sentry-systemets sensorer upptäcker rörelse larmas en operatör som kan fatta beslut om automatisk eldgivning. Samsungs talesperson Huh Kwang-hak konstaterar att medan soldater ibland somnar

eller tappar koncentrationen innebär Sentry-systemet automatisk gränskontroll utan mänskliga svagheter. Dessutom saknar systemet rädsla eller fruktan för fiendeattacker och Kwang-hak menar därför att systemet kan och kommer att förhindra krig. (Prigg, 2014)

Även vapensystemet Super aEgis II är ett automatiskt kanontorn som fungerar på liknande sätt som Sentry-systemet. När detta system upptäcker en inkräktare anropas en operatör samtidigt som systemet varnar inkräktaren om att systemet kommer att öppna eld om inte denne vänder tillbaka. Från början designades Super aEgis II-systemet för helautomatisk eldgivning, men tillverkaren har efter önskemål från kunderna implementerat olika säkerhetsspärrar, exempelvis måste en operatör fatta ett beslut om att låsa upp eldgivningsfunktionen genom att mata in ett lösenord (Parkin, 2015).

Ett annat exempel på ett autonomt vapensystem är israeliska Harop, som själv kan upptäcka och bekämpa fientliga luftvärnspositioner och ballistiska robotar. Systemet är utrustat med olika sensorer för att exempelvis upptäcka fientlig radar, och när det gjort det kan det självt bekämpa radarn (IAI, 2013).

Men utvecklingen av autonoma vapensystem berör inte bara stater, utan via internet kan man idag köpa tekniken för autonoma maskingevär (Realsentrygun.com, 2014). Med andra ord kan det finnas betydande risker för att denna typ av system även kan spridas till ickestatliga aktörer, som exempelvis para-militära grupper och terrororganisationer.

Parallellt med utvecklingen av olika nya AI-baserade vapenssystem pågår även en diskussion om hur AI kan förstärka de befintliga systemen. Inom USAs försvarsdepartement diskuteras hur de autonoma vapensystemen kan göras säkrare, och departementet konstaterar att det sannolikt kommer att vara möjligt att utrusta dessa vapensystem med kärnvapen innan lämpliga säkerhetsspärrar är fullt utvecklade och implementerade (United States Air Force, 2009, s.41):

> Authorizing a machine to make lethal combat decisions is contingent upon political and military leaders resolving legal and ethical questions. These include the appropriateness of machines having this ability, under what circumstances it should be employed, where responsibility for mistakes lies and what limitations should be placed upon the autonomy of such systems. The guidance for certain mission such as nuclear strike may be technically feasible before UAS safeguards are developed. […] Ethical discussions and policy decisions must take place in the near term in order to guide the development of future UAS capabilities, rather than allowing the development to take its own path apart from this critical guidance.

3.3 Slutsatser av tillämpningarna

Baserat på beskrivningarna ovan av de olika tillämpningarna ovan så menar jag att det finns ett antal gemensamma drag mellan de olika områdena, som jag här sammanfattar i tre avsnitt.

3.3.1 Exekutiva och assisterande tillämpningar

Utifrån beskrivningarna av de olika tillämpningarna menar jag att man kan se två huvudtyper av beslutssituationer där man med framgång använt sig av artificiell intelligens som beslutsmetod. Dels i situationer där datorns förmåga att snabbt och med hög precision analysera stora datamängder ger den ett övertag över den mänskliga hjärnan, exempelvis inom ekonomisk analys, medicinsk diagnostik eller planering av stora logistiska projekt. Dels i situationer där det mänskliga beslutsfattandet har påtagliga svagheter; exempelvis i handeln med värdepapper – där vi är relativt långsamma, i trafiken – där vi lätt tappar koncentrationen eller på slagfältet – som är stressande då det kan innebära livsfara för soldaterna. Fortsättningsvis kommer jag att benämna dessa generella tillämpningar som assisterande respektive exekutiv artificiell intelligens. Av dessa tycks de tillämpningar

som är exekutiva vara de mest kontroversiella, exempelvis automatiska handelsapplikationer, självkörande fordon och autonoma vapensystem, eftersom dessa kan sägas ersätta mänskligt beslutsfattande.

3.3.2 Ökad användning av komplexa metoder baserade på neurala nätverk

Ett annat gemensamt drag i beskrivningarna av dagens tillämpningar av AI som beslutsmetod tycks vara den utbredda användningen av neurala nätverk, särskilt då av den typ som kallas Deep learning, som kan sägas vara neurala nätverk hierarkiskt ordnat i flera olika lager (därav användningen av ordet Deep). Beskrivningen av de olika tillämpningarna visade att denna typ av artificiell intelligens börjar spridas inom exempelvis ekonomiska tillämpningar och autonoma fordon.

Denna teknik tycks vara mycket kraftfull i jämförelse med exempelvis symbolistiska ansatser, samtidigt som de beslutsalgoritmer som genereras av neurala nätverk, och särskilt då i flera led som i Deep learning, kan vara mycket svåra att överblicka och utvärdera. Jag menar därför att denna utveckling sannolikt ställer stora krav på utveckling av nya metoder för att kvalitetssäkra beslutsalgoritmerna.

3.3.3 Begränsad användning av artificiell generell intelligens (AGI)

Ytterligare ett gemensamt drag bland beskrivningarna av dagens tillämpningar av AI som beslutsmetod är den mycket begränsade användningen hittills av artificiell generell intelligens, eller AGI. Samtidigt menar jag att det finns goda skäl att anta att detta inom kort kommer att förändras av flera olika skäl.

För det första tycks tekniken vara mycket kraftfull, något som bland annat visat sig genom framgångarna för DeepMind i brädspelet go, men också genom att deras system på kort tid lärt sig spela ett mycket stort antal Atari-dataspel på en övermänsklig nivå, visat sig vara överlägsen andra tekniker vid språkanalys och språkgenerering samt att tekniken gjort det möjligt för Google att göra betydande besparingar i driften av deras datorhallar. Detta gör att tekniken tycks vara mycket kraftfull.

En annan aspekt som jag menar talar för att tekniken snart kan få stor spridning är dess mångsidighet, som bland annat illustrerades av framgångsexemplen. Genom att systemet till sin natur är generellt så kan det relativt enkelt anpassas till nya tillämpningar, till skillnad från mer specialiserade system. Detta gör att man kan anta att utvecklingen av generell artificiell intelligens sannolikt kommer att vara något av ett genombrott för användningen av artificiell intelligens, en uppfattning som även delas av exempelvis Oxford-professorn Nick Bostrom (Larsson, 2016):

> En ny maskin eller pryl kan låta oss göra mer av en specifik sak. Men generell artificiell intelligens ersätter mänskligt tänkande. Därför är det potentiellt mycket mer betydelsefullt än nästan något annat vi kan förändra på planeten.

En tredje aspekt som jag menar kommer att öka spridningshastigheten är de relativt låga trösklarna. Delar av den bakomliggande koden finns tillgänglig på internet, exempelvis finns den kod som användes av DeepMind för att skapa programmet som sedan lärde sig spela Atari-spelet Breakout tillgänglig på GitHub (Kuz, 2016), samtidigt som det räcker med en grafikprocessor (Nvidia GTX 970) som finns allmänt tillgänglig i datorhandeln och som idag kostar under 3000 kr för att driva programmet.

3.4 Tekniska begränsningar med AI som beslutsmetod

Hittills har utvecklingen inom artificiell intelligens till stor del drivits på av en kombination av tekniska och kunskapsmässiga framsteg. Genom ständiga framsteg inom både tillverkningsteknik och datorarkitektur så har datorerna blivit allt kraftfullare, samtidigt som det utvecklats allt bättre modeller för implementering av artificiell intelligens och uppfunnits nya metoder för effektivare utbildning av

algoritmerna. Men eftersom traditionella datorer inte är optimerade för maskininlärning och metoderna för både implementering och maskininlärning fortfarande är relativt nya och omogna, så finns det en betydande utvecklingspotential. Detta illustreras exempelvis av Googles Tensor Processing Unit (TPU), som är specialutvecklad för maskininlärning och som företaget menar motsvarar ett tekniksprång ekvivalent med 3 generationers (dvs 6-7 år) datorutveckling av traditionella datorprocessorer (Osborne, 2016).

Vidare finns det betydande framsteg kvar att göra inom både implementeringen av artificiell intelligens och av de metoder som används för maskininlärning. När det exempelvis gäller den generella artificiella intelligens som Deep Mind använt för att spela dataspel som Breakout nämner utvecklarna Volodymyr Mnih och Koray Kavukcuoglu att systemet lär sig utifrån att slumpmässigt utföra olika kommandon och sedan lär sig med utgångspunkt från den feedback som systemet får i form av poäng. Systemet är därför beroende av att en slumpmässig kombination av kommandon ska kunna leda till att det uppnår poäng och därmed får något att utgå från, och om det inte finns några ”enkla” poäng så kommer inte systemet någonstans. Vidare nämner utvecklarna att systemet saknar en specifik minnesfunktion och därför inte kan relatera till tidigare lösningar till specifika problem, utan bara till vad systemet lärt sig generellt. (Gibney, 2015, 3:38)

Ytterligare en begränsning är att system som baseras på komplexa ickelinjära metoder kräver mycket stora datamängder för att kunna skapa en utvecklad och kompetent beslutsmodell. AlphaGo utvecklades genom att systemet först fick analysera 160 000 mänskliga partier, och sedan utifrån detta spela ungefär en miljon partier om dagen mot sig själv för att förbättra beslutsmodellen. Så när systemet mötte Lee Sedol hade systemet erfarenhet från flera hundra miljoner partier, medan Lee kanske spelat ca 50 000 partier hela sitt liv. Men Lee kunde ändå vinna ett av de fem partierna, vilket kan ses som en bedrift med tanke på skillnaden i erfarenhet. (Lake *et al*., 2016, s.38)

Vidare kan det också konstateras att framstegen för artificiell intelligens inom exempelvis områden som kräver mer kreativa eller empatiska förmågor ännu så länge tycks vara relativt begränsade. Google har exempelvis provat att förvränga bilder genom neurala nätverk, något som visserligen skapar bilder som kan sägas vara ”intressanta” men kanske inte revolutionerande (Mordvintsev *et al.*, 2015).

Dagens mest avancerade system för artificiell intelligens har med andra ord fortfarande betydande svagheter jämfört med mänsklig intelligens, bland annat eftersom systemet saknar empatisk eller kreativ förmåga samt en generell kunskapsbank från tidigare erfarenheter, som kan ge uppslag för nya lösningar snabbare. För att kompensera för detta kan man ibland använda sig av kraftfulla ickelinjära metoder, som kräver stora mängder data för att fungera bra. Detta är tekniskt möjligt på grund av datorers förmåga att snabbt analysera stora mängder data, men som inte alltid är praktiskt eftersom det exempelvis kan vara svårt att uppnå den mängd data som krävs från relevanta situationer.

4 Sannolik teknisk AI-utveckling

En genomgående lärdom i detta arbete är svårigheterna att förutse utvecklingen inom artificiell intelligens. Som tidigare nämndes berättade Googles medgrundare Sergey Brin (WEF, 2017) att han länge var mycket tveksam till potentialen med artificiell intelligens och Armstrong & Sotala (2015) har exempelvis visat hur fel prognoserna och förutsägelserna hittills har varit. Därav kan man kanske dra slutsatsen att det inte går att förhålla sig vetenskapligt kring den framtida utvecklingen av artificiell intelligens, att vad man än kommer fram till så är det med största sannolikhet felaktigt. Men jag menar att det ändå kan göras några rimliga antaganden om den tekniska utvecklingen av artificiell intelligens på både kort och lång sikt, baserat på den generella utvecklingen inom relaterade områden som exempelvis hårdvaruutveckling.

4.1 På kortare sikt

I ett kortare tidsperspektiv menar jag att utvecklingen inom artificiell intelligens sannolikt även fortsättningsvis kommer att utvecklas i snabb takt, drivet av framsteg inom hårdvaruutveckling, nya hårdvarudesigner anpassade för artificiell intelligens samt att man sannolikt kommer att finna metoder som överkommer de hinder som finns för utvecklingen, varav några nämndes under avsnittet tekniska begränsningar. Exempelvis kan den kraftiga ökningen av patent relaterade till artificiell intelligens ses som en indikation på den framtida utvecklingen.

Inom hårdvaruutvecklingen menar jag att det är rimligt att anta att den exponentiella utveckling av generell datorkraft som beskrivs av Moores lag kommer att fortsätta åtminstone några år till (Intel, 2016, Pratt, 2015, s.53). Detta kommer att medföra ytterligare processorkraft, som kan användas för kraftfullare AI-tillämpningar.

Vidare menar jag att det kan antas att det finns ett betydande utrymme för prestandaförbättringar genom att utveckla nya specialiserade kretsar liknande Googles TPU (Osborne, 2016). Paralleller kan sannolikt dras till exempelvis utvecklingen inom brytningen av kryptovalutan Bitcoin, som effektiviserats betydligt från det att man använde först vanliga processorer för att utföra de beräkningar som ligger bakom valutan. I ett första steg gick man över till att använda grafikprocessorer för att sedan börja använda specialiserade kretsar, s.k. ASICs (Bitcoin.se, 2013). Som jag ser det talar mycket för att hårdvaran bakom artificiell intelligens sannolikt kommer att göra en liknande utveckling.

När det gäller mjukvarusidan så finns det även där utrymme för betydande förbättringar. Med tanke på de insatser som görs för att förbättra artificiell intelligens anser jag därför att det sannolikt att kommer att utvecklas allt bättre metoder för snabbare och bättre inlärning, metoder som exempelvis inte kräver lika stora mängder data och metoder för att överföra någon form av ”förförståelse” som låter systemet få en grundförståelse för tillvaron, som det sedan kan utgå ifrån. Själva minneshanteringen kommer säkert också att förbättras, exempelvis så att AI-systemet en dag kanske får mer av ett episodiskt minne och därför kan relatera till specifika situationer med utgångspunkt i tidigare erfarenheter.

Vidare menar jag att utvecklingen av artificiell generell intelligens (AGI) sannolikt också kommer påskynda utvecklingen inom artificiell intelligens eftersom systemen inte längre behöver byggas speciellt för varje specifik lösning, utan AGI-system kan designas som ”stapelvara” och sedan anpassas via maskininlärning.

En tänkbar lite mer avlägsen utvecklingsväg för en förbättrad maskininlärning skulle även potentiellt kunna vara att ge systemet någon form av dedikerade mekanismer som motsvarar den biologiska hjärnans belöningssystem. Som nämndes i inledningen kan man argumentera för att just kopplingen mellan belöningssystemen och inlärning varit avgörande för utvecklingen från enklare levande organismer till individer med upplevelser och medvetande (Ginsburg & Jablonska, 2007), och därmed också blivit effektivare på inlärning. Visserligen bygger redan dagens AI-program ofta på att systemet i någon mening försöker nå en slags maxpoäng givet vissa förutsättningar, men belöningssystemet kan sannolikt både förstärkas, bli mer generellt och även, precis som hjärnan, ta hjälp av flera olika inputs/sensorer/sinnen.

4.2 På lite längre sikt

I ett längre tidsperspektiv finns det många som menar att artificiell intelligens en dag kommer att bli mer kraftfull än den mänskliga hjärnan, drivet av tekniska framsteg och förbättrade metoder. Exempelvis menar filosofen Nick Bostrom att en generell artificiell intelligens skulle vara den mest betydelsefulla skapelsen för mänskligheten någonsin och den kan bli ett hot mot den mänskliga artens överlevnad. Enligt honom kommer generell artificiell intelligens antingen öka välståndet på planeten gränslöst, eller döma oss till undergång (Larsson, 2016).

Resonemanget bygger på att tekniken kan användas så att den hela tiden blir allt bättre och mer intelligent. Men den skulle också bli bättre på att träna sig själv, så att förbättringstakten kommer sannolikt att accelerera. Därför menar Bostrom att det är svårt att tänka sig att en artificiell intelligens skulle nå en lågbegåvad människas nivå men sedan stanna där. Sannolikt skulle den mycket snabbt – kanske inom månader, dagar eller några minuter – vara många gånger mer intelligent än den smartaste människan, en händelse som ibland kallas teknisk singularitet. Som tidigare nämnts menar exempelvis Googles teknikchef Ray Kurzweil att detta sannolikt inträffar ca år 2045, baserat på hans teorier om den generella teknikutvecklingen (Larsson, 2016, Kurzweil, 2005, s.120).

Antagandet om en i det närmaste exponentiell utveckling av intelligens och kapacitet stämmer exempelvis väl med Hassabis (2016a) observation av utvecklingen av AlphaGo. Enligt honom tog det flera år att utveckla ai-system för att nå en duktig amatörs nivå. Men när man väl nådde dit så tog det bara ungefär ett halvår innan man kunde slå den bästa människan i världen.

5 Tänkbara samhällskonsekvenser av AI-utvecklingen

Utvecklingen inom artificiell intelligens går mycket fort samtidigt som den potentiella samhällspåverkan är mycket stor. Därför menar jag att inget generellt arbete om artificiell intelligens inom besluts-, risk- och policyanalys är komplett utan en diskussion om risker och möjligheter kopplade till utvecklingen inom området. Samtidigt är det knappast möjligt att i denna typ av arbete diskutera alla aspekter och jag har därför valt att här fokusera på samhällskonsekvenserna ur två specifika frågeställningar; vilken påverkan kan en ökad användning av artificiell intelligens få för arbetsmarknad och sysselsättning samt hur kan artificiell intelligens förändra de internationella relationerna.

5.1 Arbetsmarknaden

En aspekt som diskuterats ganska mycket i samband med artificiell intelligens är dess påverkan på ekonomin i allmänhet, och särskilt på arbetsmarknaden. Tillämpningarna av artificiell intelligens lovar visserligen effektivare produktion och därmed billigare produkter, men samtidigt är det många som oroar sig för konsekvenserna för att tekniken kan leda till lägre sysselsättning när robotar och andra tillämpningar av artificiell intelligens i snabb takt tar över allt fler av de arbetsuppgifter som människan utför idag. Exempelvis säger författaren Martin Ford (Olsson Jeffrey, 2016):

> Maskiner börjar tänka och få kognitiva förmågor. Det här handlar om att inte bara ersätta muskelkraft, utan hjärnkraft. Vad kan genomsnittsmänniskan egentligen göra? De flesta gör saker som är rutinmässiga och förutsägbara, och de jobben kommer att vara mottagliga för automation.

Samtidigt framhåller förespråkarna av utvecklingen ofta att marknadsekonomin genom historien visat sig mycket motståndskraftig mot förändring. Man menar att ekonomin och arbetsmarknaden genomgått betydande förändringar i samband med exempelvis den industriella revolutionen och att dessa förändringar hela tiden lett till högre välstånd. Vidare framhåller man ofta att många av dagens arbetstillfällen, kopplade till exempelvis internet och sociala media, inte ens existerade för 15 år sedan. (Se t.ex. Autor, 2015, Kurzweil, 2016, 10:00)

Vid en analys av argumenten från de olika sidorna kan man se tydliga paralleller mellan de argument som de som oroar sig för AI-utvecklingens konsekvenser och Michael Porters (1979) marknadsmodell med fem marknadskrafter. Enligt denna modell beror lönsamheten på en marknad på konkurrensen inom sektorn, från substitut, från nya aktörer samt från leverantörernas och köparnas relativa styrka. Om man betraktar hela den mänskliga globala arbetsmarknaden som en marknad, så kan man se artificiell intelligens som ett substitut till mänskliga arbetstjänster. I enlighet

med Porters modell borde därmed artificiell intelligens kunna minska lönsamheten för mänskligt arbete, under förutsättning att inte den totala efterfrågan ökar.

Å andra sidan så har de som menar att utvecklingens konsekvenser kommer att vara förhållandevis begränsade onekligen en poäng i att exempelvis den industrialiserade världens ekonomi och arbetsmarknad även tidigare genomgått flera skiften och varje gång repat sig, rimligen eftersom ekonomin tycks ha en latent förmåga att skapa nya arbetstillfällen. Detta kan även sägas vara i linje med exempelvis Acemoglu & Autor (2011) och Autor (2015), och något som knappast ändras i och med att artificiell intelligens utför allt mer arbete.

Med tanke på teknikens kraftfullhet och osäkerheten kring utvecklingen har det på senare tid publicerats flera studier som försöker förutspå automatiseringens och den artificiella intelligensens påverkan på arbetsmarknaden. Frey och Osborne publicerade 2013 en uppmärksammad undersökning, som kom fram till att 47 % av arbetstillfällena i USA var i farozonen att ersättas av automatisering, och när arbetsmarknaden i Sverige analyserades med samma ansats kom man fram till att motsvarande siffra i Sverige är 53 % (Fölster, 2014). Som svar på denna studie gjorde forskare vid OECD 2016 en analys där författarna identifierade att i genomsnitt 9 % av arbetstillfällena inom OECD (exempelvis 9 % i USA och ca 7 % i Sverige) är i farozonen att ersättas med automatisering och artificiell intelligens (Arntz *et al*., 2016). Med andra ord tycks det finnas en betydande osäkerhet kring hur den tekniska utvecklingen med artificiell intelligens kan tänkas påverka sysselsättningen.

Att i detta arbete ingående granska de olika studierna för att försöka utröna vilken som är mest trovärdig ligger utanför detta arbetes gränser, men baserat på tidigare slutsatser om artificiell intelligens menar jag att flera av slutsatserna i Arntz *et al.* (2016) kan ifrågasättas. Exempelvis skriver Arntz *et al* (2016, s.4) i sina slutsatser att de anser sig ha identifierat tre orsaker till att utvecklingen sannolikt kommer att ha en begränsad påverkan på arbetsmarknaden:

> First, the utilisation of new technologies is a slow process, due to economic, legal and societal hurdles, so that technological substitution often does not take place as expected. Second, even if new technologies are introduced, workers can adjust to changing technological endowments by switching tasks, thus preventing technological unemployment. Third, technological change also generates additional jobs through demand for new technologies and through higher competitiveness.

Den första slutsatsen, att införandet av tekniker som artificiell intelligens och automatisering är en långsam process, menar jag motsägs av den generella tekniska utvecklingen inom området och alla observationer av hur många industrier och fabriker automatiseras i mycket snabb takt.

Ett exempel är taiwanesiska Foxconn, som tycks ha kommit mycket långt i sin plan att övergå till i den närmaste helt automatiserad produktion. Företaget har fattat ett beslut om ett trestegsprogram för automatiserade fabriker, där det första steget innebär att enskilda robotar används för att automatisera farliga, svåra eller tråkiga arbetsuppgifter. Det andra steget innebär att hela produktionslinjer automatiseras och det tredje steget innebär att hela fabriker automatiseras, med ett minimalt antal anställda som övervakar produktion, logistik och kvalitetstestning. Enligt Foxconns automatiseringschef Dai Jia-peng har flera av företagets anläggningar nu nått steg två eller tre och företaget har hittills implementerat 40 000 industrirobotar i produktionen, med en ökningstakt på ca 10 000 robotar per år (dvs mer än en ny robot i timmen dygnet runt, alla dagar året runt). Detta har bland annat medfört att en av Foxconns fabriker har kunna minska antalet anställda från 110 000 till 50 000, och eftersom beroendet av billig personal nu minskar samtidigt som världen tycks bli allt mer protektionistisk så överväger företaget att börja flytta produktion till exempelvis USA. (Lee & Hwang, 2016, Mölne, 2016, Wu, 2017)

Andra exempel är Shenzhen Evenwin Precision Technology, som maj 2015 meddelade att de nu öppnar en helt automatiserad fabrik som ersätter 90 % av företagets anställda samt elbiltillverkaren Teslas planer på helt automatiserad

produktion i USA av storsatsning model 3 (Cruickshank, 2015, Fung, 2016). Med hänsyn till dessa och andra liknande rapporter så menar jag att slutsatsen att införandet av tekniker som artificiell intelligens och automatisering är en långsam process kan ifrågasättas.

Man kan visserligen argumentera att detta framförallt gäller enklare industriproduktion som redan idag framförallt är lokaliserad till låglöneländer som Kina, men exempelvis Teslas planer för en helautomatisk fabrik i USA tyder som jag ser det på att tekniken snart kommer att sprida sig även till mer avancerad produktion inom exempelvis fordonsindustrin, något som därmed säkert påverkar stora arbetsgivare som Volkswagen, Mercedes, GM, Ford, BMW, Volvo, Renault, Scania med flera. Företag som alla är globala och därför möter en global konkurrens, och därför hela tiden måste anpassa sin produktion för att behålla sin konkurrenskraft, samtidigt som de fortfarande har betydande produktion i exempelvis Europa eller USA.

Även den andra och den tredje slutsatsen i Arntz *et al.* (2016) framstår som något tveksamma. Att arbetare enkelt skulle kunna behålla arbetet genom att bara skifta arbetsuppgifter sägs emot av de stora personalneddragningarna, i de nämnda exemplen mellan drygt hälften till nio av tio, och i den mån den nya tekniken leder till nya arbeten inom exempelvis artificiell intelligens så kräver dessa uppgifter sannolikt en helt annan kompetensprofil än de arbetstillfällen som ersätts.

Sammantaget menar jag därför att man sannolikt bör vara mycket försiktig i bedömningen av hur och när arbetsmarknaden påverkas av automatisering och artificiell intelligens, men att det finns betydande indikationer på att AI och automatisering snabbt kan få ett betydande genomslag. Exempelvis skriver Pratt (2015):

> Robots are already making large strides in their abilities, but as the generalizable knowledge representation problem is addressed, the growth of robot capabilities will begin in earnest, and it will likely be explosive. The effects on economic output and human workers are certain to be profound.

och i en rapport till president Barack Obama konstaterades (Furman *et al.*, 2016, s.14):

> If these estimates of threatened jobs translate into job displacement, millions of Americans will have their livelihoods significantly altered and potentially face considerable economic challenges in the short- and medium-term.

För att skapa sig en generell bild över hur arbetsmarknaden kan påverkas av artificiell intelligens och automatisering menar jag att man sannolikt kan dra vissa paralleller till de effekter som följt med den s.k. globaliseringen, där en stor del av västvärldens produktion och produktionstillfällen flyttat till låglöneländer. I båda fallen har det med andra ord handlat om en form av substitut, även om det sannolikt också finns skillnader i vilka grupper och arbeten som berörs, något som stör jämförelsen. I samband med globaliseringen har många uppmärksammat att arbetstagarnas del av den totala ekonomin tycks ha minskat i många industrialiserade länder. Enligt IMF har exempelvis lönerna i USA minskat från knappt 65 % till lite drygt 60 % som del av den totala ekonomin under perioden 1980-2005, och från ca 73 % till ca 64 % i Västeuropa under samma period (Jaumotte & Tytell, 2007, Figure 8, s.39).

När IMF analyserar de bakomliggande orsakerna kommer man dock fram till att det framförallt är stora teknikskiften inom informations- och kommunikations-sektorerna som ligger bakom denna förändring, inte primärt att arbetstillfällen flyttats utomlands (Jaumotte & Tytell, 2007, s.19). Samtidigt kan man fråga sig om inte artificiell intelligens kan ses som just ett stort teknikskifte inom informations- och kommunikationssektorerna, fast potentiellt betydligt mer kraftfullt är de tidigare teknikerna och därmed fundera på vad detta kan tänkas få för påverkan på lönesättning och ekonomiska klyftor. Med tanke på den stora osäkerheten inom området finns det med andra ord sannolikt mycket goda skäl för att beslutsfattare att noggrant bevaka

utvecklingen och utarbeta planer för att förebygga och vid behov kunna möta en snabb förändring på arbetsmarknaden, både när det gäller löner och sysselsättning.

5.2 Internationella relationer

Även ur ett statsvetenskapligt perspektiv menar jag att det finns skäl att reflektera över vilka konsekvenser som kan följa med en ökad användning av artificiell intelligens, både när användningen av AI ökar inom militären och hur den ekonomiska dynamiken förändras när artificiell intelligens får allt större påverkan inom produktion och ekonomi.

Inom statsvetenskapen finns många olika synsätt på internationella relationer, som har olika förklaringsmodeller för varför stater agerar som de gör. Exempelvis utgår realismen från att världen präglas av anarki och att staternas främsta drivkraft därför är att hela tiden garantera sin säkerhet, något som därför bland annat minskar utrymmet för mellanstatligt samarbete (Mowle, 2003, s.561). I enlighet med detta synsätt söker stater hela tiden makt och inflytande (hegemoni) över varandra, ytterst baserat på militär kapacitet. Exempelvis skriver realisten Mearsheimer (2001, s. 29):

> Great powers, I argue, are always searching for opportunities to gain power over their rivals, with hegemony as their final goal. This perspective does not allow for status quo powers, except for the unusual state that achieves preponderance. Instead, the system is populated with great powers that have revisionist intentions at their core.

Det andra stora perspektivet inom statsvetenskap, liberalismen, utgår istället från att stater kan tillvarata sina intressen genom att utveckla både formellt och informellt samarbete med andra stater. Realismen och liberalismen har med andra ord mycket olika synsätt på drivkrafterna för staters agerande och förutsättningarna för internationellt samarbete, och det har därför under lång tid varit en intensiv akademisk diskussion om vilket synsätt som har störst förklaringsvärde (Mowle, 2003, s.561).

Inom den internationella politiken finns det gott om exempel på både situationer där stater agerat på sätt som bäst förklaras av realismen och på situationer där stater agerat med utgångspunkt i liberalismen. Med andra ord tycks det inte vara så enkelt att ett synsätt, liberalism eller realism, alltid förklarar staternas agerande utan att staters agerande snarare tycks vara avhängigt av faktorer i den aktuella situationen (Mowle, 2003, s.584):

> This research indicates that the Western democracies included in the sample express problem representations consistent with liberalism and realism under the situational and systemic circumstances predicted in international relations literature. Such states will be more likely to express problem representations consistent with a liberal worldview when their security is guaranteed by another power, regardless of the overall distribution of power in the system. These states are also more likely to express problem representations consistent with a realist worldview when a rival, ally, or fellow democracy is involved in the conflict.

En som jag ser det viktig slutsats i Mowles (2003) arbete är att stater sannolikt är mer benägna att söka samarbete när deras säkerhet garanteras, medan det vid större säkerhetspolitisk osäkerhet finns en större benägenhet för maktspråk i enlighet med realismen. Detta menar jag talar för att om exempelvis utvecklingen leder till säkerhetspolitiska obalanser till följd av ensidig utveckling och användning autonoma vapensystem så kan man med detta synsätt sannolikt förvänta sig en större benägenhet för kapprustning och andra maktyttringar, för i enlighet med realismen utvidga sin maktbas och på så sätt försöka säkerställa den nationella säkerheten. En sådan utveckling skulle därmed också minska utrymmet för internationellt samarbete och överenskommelser.

Även FN ser påtagliga faror med utvecklingen av autonoma vapensystem. I en utredning av utvecklingen av Lethal autonomous robotic (LAR) konstateras att autonoma vapensystem riskerar att skapa betydande säkerhetspolitiska obalanser,

minska trösklarna för både externa och interna konflikter, medföra ökad kapprustning samt potentiellt även medföra att denna typ av teknik även blir tillgänglig för icke-statliga grupperingar, som exempelvis kan antas vara mindre benägna att följa internationella avtal och konventioner. (Heyns, 2013. s.16)

Sammantaget väljer FN-rapporten därför att rekommendera ett förbud av autonoma vapensystem samt att man skapar ett utskott för att arbeta fram gemensamma internationella riktlinjer för autonoma vapensystem (Heyns, 2013. s.1):

> The Special Rapporteur recommends that States establish national moratoria on aspects of LARs, and calls for the establishment of a high level panel on LARs to articulate a policy for the international community on the issue.

Men även om FNs rekommendationer kan ses som rimliga ur ett statsvetenskapligt perspektiv, så kan man fundera över deras realpolitiska realism. Till skillnad från exempelvis atomvapen, som exempelvis kräver stora investeringar i kärnreaktorer och centrifuger så menar jag att den bakomliggande tekniken för autonoma vapensystem är mycket lättare att tillämpa för alla med tillräckliga kunskaper. Är det exempelvis ens praktiskt möjligt att förhindra utvecklingen av autonoma vapensystem, när tekniken i sig är så tillgänglig att det enkelt och till en låg kostnad går att beställa tekniken som behövs för att skapa ett autonomt maskingevär via internet (Realsentrygun.com, 2014).

En annan aspekt att reflektera över när man funderar över artificiell intelligens och automatisering i ett internationellt perspektiv är hur tekniken kan påverka förutsättningarna för de stater som idag ännu inte utvecklat en stor industrisektor att utveckla välstånd, eller om tekniken kanske riskerar att skapa permanenta klyftor mellan de industrialiserade länderna och resten av världen. Traditionellt sett har länder kunnat utvecklas genom att initialt konkurrera internationellt med låga löner inom enklare tillverkningsindustri för att därigenom bygga upp sin industriproduktion, för att sedan successivt höja produkternas kvalitet och komplexitet, och därigenom kunna ta bättre betalt och höja sin levnadsstandard. Med artificiell intelligens så ligger det nära till hands att anta att det blir mycket svårare för dessa länder att ta det första viktiga steget och konkurrera med låga löner inom enklare tillverkning.

Slutligen menar jag att det kan vara väl värt att stanna upp och reflektera kring de miljökonsekvenser som följer när artificiell intelligens används för att effektivisera industriproduktionen, och därmed bidra till ökad konsumtion. Redan nu höjs ofta starka röster för att den konsumtion som vi har idag inte är hållbar, ett problem som sannolikt kommer att öka med billigare produktion. (WWF, 2016)

6 Diskussionen om artificiell intelligens

Artificiell intelligens är redan idag en mycket kraftfull teknik, med många tillämpningar där den mänskliga intellektuella kapaciteten överträffas. Men ändå så finns det mycket som talar för att vi bara är i början av denna utveckling och att konsekvenserna på ekonomi och samhälle sannolikt kommer att eskalera i takt med att den artificiella intelligensen blir allt mer kraftfull. Därför pågår det idag en mycket intensiv akademisk debatt kring AIs konsekvenser på samhället, där utvecklingen analyseras ur exempelvis filosofiska, juridiska eller ekonomiska perspektiv.

6.1 The Future of Lifes upprop

Ett inlägg i debatten är exempelvis organisationen the Future of Life Institutes upprop för säkrare utveckling av artificiell intelligens (Future of life, 2016). Detta upprop har undertecknats av många namnkunniga AI-forskare och IT-entreprenörer, exempelvis Stuart Russell (Berkeley University), Francesca Rossi (Harvard), Nick Bostrom (Oxford University), Demis Hassabis (DeepMind), Ben Goertzel (OpenCog), Eric Horvitz (Microsoft), Peter Norvig (Google), Elon Musk (SpaceX & Tesla), Yann

LeCun (Facebook), Steve Wozniak (Apple), Erik Brynjolfsson (MIT), Stephen Hawking (Cambridge) och många många fler (Future of life, 2016).

I det tillhörande dokumentet till uppropet ”Research Priorities for Robust and Beneficial Artificial Intelligence” tar Stuart Russell, Daniel Dewey och Max Tegmark (2015) upp ett flertal aspekter av utvecklingen inom artificiell intelligens; exempelvis AIs påverkan på de ekonomiska systemen, AI ur ett filosofiskt och etiskt perspektiv, datasäkerhetmässiga aspekter av utvecklingen av artificiell intelligens och mer långsiktiga frågeställningar relaterade till den dag som artificiell intelligens kapacitet överstiger den mänskliga hjärnan. Syftet med uppropet är enligt initiativtagarna inte att hindra utvecklingen av artificiell intelligens utan en uppmaning till säker och ansvarstagande utveckling av tekniken (Russell, Dewey, Tegmark, 2016):

> In summary, success in the quest for artificial intelligence has the potential to bring unprecedented benefits to humanity, and it is therefore worthwhile to research how to maximize these benefits while avoiding potential pitfalls. The research agenda outlined in this paper, and the concerns that motivate it, have been called ‘‘anti-AI”, but we vigorously contest this characterization. It seems self-evident that the growing capabilities of AI are leading to an increased potential for impact on human society. It is the duty of AI researchers to ensure that the future impact is beneficial. We believe that this is possible, and hope that this research agenda provides a helpful step in the right direction.

6.2 ITIF: Farorna med artificiell intelligens är överdrivna

Men det finns även de som menar att varningarna för artificiell intelligens är kraftigt överdrivna. Exempelvis utnämndes Future of Life Institutes upprop till årets ”Luddite”, med hänvisning till de som under det tidiga 1800-talet försökte motarbeta den industriella utvecklingen inom textilindustrin genom att slå sönder vävstolar, av den amerikanska tankesmedjan och lobbyorganisationen Information Technology & Innovation Foundation (ITIF, 2016a).

Enligt ITIF är de farhågor och varningar som Future of Life Institute väcker i sitt upprop kraftigt överdrivna, både när det gäller utvecklingen generellt inom artificiell intelligens och när det gäller automatiserade vapensystem med artificiell intelligens. När det gäller frågorna kring artificiell generell intelligens argumenterar ITIF särskilt mot Nick Bostroms tankar, och skriver (Atkinson, 2015):

> In his book Superintelligence: Paths, Dangers, Strategies, Oxford professor Nick Bostrom reflected the general fear that “superintelligence” in machines could outperform “the best human minds in every field, including scientific creativity, general wisdom and social skills.” He argues that artificial intelligence will advance to a point where its goals are no longer compatible with that of humans and, as a result, superintelligent machines will seek to enslave or exterminate us.
>
> […]
>
> Raising such sci-fi doomsday scenarios just makes it harder for the public, policymakers, and scientists to support more funding for AI research. Indeed, continuing the negative campaign against artificial intelligence could potentially dry up funding for AI research, other than money for how to control, rather than enable AI. What legislator wants to be known as “the godfather of the technology that destroyed the human race”? (On the other hand, if we are all dead, then one’s reputation is the last of one’s worries.)

Budskapet från ITIF tycks med andra ord vara att den kritiska debatten kring artificiell intelligens är mycket skadlig eftersom den hotar forskningsanslagen och riskerar att öka lagkraven kring artificiell intelligens. Istället menar ITIF att artificiell intelligens är en teknik med mycket stor potential och som bara är i början av sin utveckling, bara vi inte låter oss hindras av en paranoid rädsla.

ITIF har även mycket starka åsikter om debatten om att förbjuda autonoma vapensystem (Atkinson, 2015):

> As part of the paranoia over artificial intelligence, efforts to establish a global ban on offensive autonomous weapons—also known as "killer robots"—have intensified.
>
> […]
>
> But that overlooks the fact that the military clearly will benefit, because substituting robots for soldiers on the battlefield will increase a military's capabilities while substantially decreasing the risk to its personnel. Furthermore, it is possible that autonomous weapons could be programmed to engage only known enemy combatants, which may even lead to a reduction in civilian casualties.
>
> Moreover, military research and investment has long been a key catalyst to developing and commercializing new technologies with important commercial uses, and robotics will likely prove no different.
>
> […]
>
> Unfortunately, the efforts to ban autonomous weapons could very well reduce funding and support for robotics research that would have significant positive spillovers for other kinds of robotics use that would in turn increase economic productivity and quality of life over the next half-century. With autonomous robots, factories will be able to increase productivity and better compete with low-cost competitors, mines will be able to improve safety, and hospitals will be able to provide better care to patients. Substituting robots for human workers will lead to higher productivity, lower costs, higher wages, more capabilities, and more service availability, all without reducing the total number of jobs as demand expands across the economy in response to increasing supply.
>
> Indeed, robotics could very well be the most important technology of the 21st century. As such, the battle to ban autonomous weapons, much like the fight over artificial intelligence, works against the societal goal of building innovations that will improve human lives. Rather than allowing those predicting doom and gloom to dominate the debate, policymakers should encourage military investment in autonomous robots, not only to improve national defense, but also to accelerate the development of autonomous robots for other sectors.

ITIF argumenterar med andra ord för att autonoma vapensystem bör tillåtas eftersom de stärker militären och minskar risken för soldaterna. Enligt ITIF kan autonoma vapensystem potentiellt också programmeras för att endast bekämpa fiender och inte civilbefolkningen. Vidare menar ITIF att forskningen inom autonoma vapensystem är viktig eftersom framsteg i militär utveckling ofta leder till tekniker som även har civilt värde.

6.3 Analys av diskussionen

Med tanke på det höga tonläget i inlägget från Information Technology & Innovation Foundation är det lätt att få intrycket att parterna i debatten står mycket långt ifrån varandra, men jag menar att de i grund och botten sannolikt har ganska mycket gemensamt. Exempelvis ger båda inläggen uttryck för att artificiell intelligens är en mycket kraftfull teknik, till skillnad från de som menar att artificiell intelligens är en teknik som alla andra.

Båda sidor är också eniga om att tekniken har mycket stor potential att påverka i stort sett alla dimensioner av det mänskliga samhället, till skillnad från de som vill tona ner betydelsen av artificiell intelligens. Båda sidor är även eniga om att ekonomin och sysselsättningen kan påverkas genom att arbetsuppgifter som tidigare utförts av människor i stället utförs effektivare av artificiell intelligens och robotar, och att säkerhetspolitiken kan påverkas genom att även militären blir effektivare samtidigt som färre soldaters liv riskeras.

Samtidigt kommer ITIF och The Future of Life Institute till helt olika slutsatser, där de förstnämnda vill utveckla artificiell intelligens så snart som möjligt och tycks utgå från att konsekvenserna av artificiell intelligens ofrånkomligt uteslutande kommer att vara positiva, i alla fall för dem. Enligt ITIF kommer en ökad användning

av artificiell intelligens inom näringslivet medföra effektivare produktion, billigare produkter, högre löner och högre välstånd samtidigt som mängden arbetstillfällen ökar som svar på den högre tillgången. För mig framstår detta som en mycket egen tillämpning av de nationalekonomiska begreppen tillgång och utbud där visserligen ett ökat utbud i sig leder till större volym, men vid en lägre prispunkt dvs i detta fall lägre löner (Eklund, 1995, s.60). Dessutom bortser ITIF från den tröghet som finns i det ekonomiska systemet, att det tar ofta tid för människor som förlorat sitt arbete i samband med exempelvis en nedläggning att ställa om och hitta ett nytt, samt att effekterna av ett inkomsttapp kan bli mycket långvariga (Furman *et al.*, 2016, s.12):

> Perhaps more significantly, over time, displaced workers' earnings recover only slowly and incompletely. Even ten or more years later, the earnings of these workers remain depressed by 10 percent or more relative to their previous wages. These results suggest that for many displaced workers there appears to be a deterioration in their ability either to match their current skills to, or retrain for, new, in-demand jobs. AI-driven automation can act—and in some cases has already acted—as a shock to local labor markets that can initiate long-standing disruptions.

Vidare tycks ITIF mena att autonoma vapensystem uteslutande är positivt, dels eftersom detta kommer att driva på den allmänna utvecklingen av artificiell intelligens ytterligare och dels eftersom man då kan utveckla effektivare robotar som bara bekämpar fiender och inte civilbefolkningen. Men redan idag satsas mycket stora pengar på utvecklingen av artificiell intelligens, något som exempelvis framgår av den mycket kraftiga ökning av patent inom området som tidigare redovisats. Som jag ser det finns det med andra ord mycket lite som talar för att utvecklingen av artificiell intelligens skulle stå still om man inte satsade mer på att utveckla militära tillämpningar.

När det gäller idén om att autonoma vapensystem skulle innebära en säkrare och effektivare militär, så finns det många tänkbara motargument och invändningar. Exempelvis kan man som tidigare konstaterats notera att egenskaper som empati och medkänsla sannolikt kommer att vara bland det svåraste att överföra till artificiell intelligens eftersom dessa egenskaper bygger på samspel mellan flera olika hjärnstrukturer, exempelvis spegelneuroner, amygdala och belöningsfunktionerna, medan vi när vi skapar artificiell intelligens använder oss av en central målfunktion för att uppnå önskat beteende. Som jag ser det är empati och medkänsla sannolikt den starkaste garanten mot omfattande övergrepp, eftersom offrets lidande gör att den normala reaktionen är att undvika beteendet då offrets smärta i normala fall blir vår egen. Detta kan exempelvis förklara att man i samband med omfattande krigsförbrytelser och folkmord ofta ser en retorik där motståndarsidan avhumaniseras för att bryta denna koppling mellan offer och förövare (Stanton, 2013).

En annan aspekt är att ökad användning av autonoma vapensystem sannolikt kommer att förändra den säkerhetspolitiska balansen. I ett tidigare avsnitt argumenterade jag till att detta i sig kan leda till kapprustningar, ökade geopolitiska spänningar och ett instabilare säkerhetsläge när länder i allt större utstäckning använder olika former av maktspråk för att försöka uppnå makt och inflytande.

Vidare har mänskligheten genom tiderna uppvisat en betydande uppfinnings-rikedom när det gäller att uppfinna destruktiva användningsområden för ny teknik. Kemiska sprängämnen som krut och dynamit var avgörande för att bygga järnvägsnäten tvärs över kontinenterna, men har också gett oss effektivare och dödligare vapen och därmed fruktansvärda krig och konflikter världen över. Zyklon B utvecklades från början som ett bekämpningsmedel för att öka skördarna och bekämpa undernäring, men är idag mest känt för användningen i gaskamrarna i Auschwitz. Elektriciteten som ger oss ljus varje dag används också för tortyr och som avrättningsmetod i den elektriska stolen. Atomkraften ger oss daglig energi till att driva kylskåp, spisar, datorer och allt annat elektiskt, men restprodukterna är giftiga i hundratusentals år och det räcker med en knapptryckning för att med kärnvapen utrota allt mänskligt liv på jorden inom loppet av någon timme. Till och med vatten, som är

ett fundament för livet på jorden och dessutom en ”ren” energikälla, används för tortyr – sk waterboarding. Tanken på att en så kraftfull teknik som artificiell intelligens enbart skulle användas på ett positivt och konstruktivt sätt framstår därför som både historielöst och naivt.

I detta ljus menar jag att den inställning som förs fram av The Future of Life Institute framstår som liberal och pragmatisk, med ett tydligt fokus på teknisk säkerhet och betoning av behovet av mer kunskap om vilka effekter som ökad användning av artificiell intelligens kan antas få för våra samhällen, exempelvis ut ett ekonomisk eller legalt perspektiv. Samtidigt efterfrågar The Future of Life Institute en fördjupad diskussion mer långsiktiga utvecklingsfrågor gällande exempelvis superintelligens och teknisk singularitet – som även om de idag framstår som mycket avlägsna med största sannolikhet en dag kommer att inträffa. Ur ett etiskt perspektiv kan man som jag ser det argumentera för att denna inställning som en miniminivå, helt oavsett om man exempelvis argumenterar ur ett utilitaristiskt eller pliktetiskt perspektiv.

7 Slutsatser

I detta arbete har många olika aspekter av artificiell intelligens som beslutsmetod behandlats, från funktionella skillnader mellan hjärnan och datorkretsar till sannolika konsekvenser för arbetsmarknad och internationell säkerhetspolitik. Baserat på dessa diskussioner sammanfattar jag här några av de viktigaste slutsatserna kring användningen av artificiell intelligens som beslutsmetod.

7.1 Artificiell intelligens som beslutsmetod

I inledningen av arbetet jämfördes hjärnans funktionssätt med datorns och baserat på detta identifierades två typsituationer där det mänskliga beslutsfattandet har påtagliga svagheter i relation till datorer, dels i komplexa situationer med stora mängder data och dels i situationer där hjärnan av olika anledningar är otillförlitlig, exempelvis på grund av att den är långsam i jämförelse med en dator eller att den har lätt bli uttröttad, distraherad eller stressad. Svagheter som skapar problem för beslutsfattandet inom många olika områden, exempelvis inom ekonomin, trafiken, medicinen eller för militären och där man därför kan tänka sig att det finns betydande potential för tekniska lösningar för bättre beslutsfattande.

I ett senare avsnitt i arbetet konstaterades att det inom alla dessa områden pågår ett intensivt och i många avseenden mycket framgångsrikt arbete med att på olika sätt tillvarata fördelarna med beslutssystem baserade på artificiell intelligens. Exempelvis nämndes AI-system som assisterar läkare med rekommendationer om olika cancerbehandlingar baserade på omfattande dataanalys av den senaste forskningen och olika system som ersätter det mänskliga beslutsfattandet, exempelvis genom autonoma fordon eller vapensystem.

Samtidigt kan det också konstateras att dagens system för artificiell intelligens fortfarande har betydande svagheter och begränsningar inom många områden. Exempelvis är tillämpningar inom områden som kräver mer kreativa eller empatiska förmågor ännu så länge relativt begränsade och man kan anta att de tekniska utmaningarna inom dessa områden är större än för exempelvis uppgifter som är rutinmässiga eller inriktade på analys av stora datamaterial.

Vidare kräver de mest kraftfulla metoderna idag stora mängder data för att ge bra resultat. Detta är tekniskt möjligt på grund av datorers förmåga att snabbt analysera stora mängder data, men som inte alltid är praktiskt genomförbart eftersom det kan vara svårt att uppnå den mängd data som krävs från relevanta situationer.

Det finns många olika tekniker och metoder för att skapa artificiell intelligens för beslutsfattande, alla med sina styrkor och svagheter. Generellt så beror valet av metod till stor del på en avvägning mellan förutsägbarhet och transparens å ena sidan och kraftfullhet å andra sidan. Exempelvis så är de traditionella symbolistiska metoderna i

regel relativt lätta att analysera och förstå, medan metoder som exempelvis bygger på flera ickelinjära metoder, s.k. Deep learning, kan vara mycket svåra. Samtidigt har de sistnämnda metoderna visat sig vara betydligt kraftfullare i många tillämpningar, varför dessa fått ett betydande genomslag på senare tid.

7.2 Utvecklingens konsekvenser

För att kunna redogöra för några av de potentiella risker som utvecklingen inom artificiell intelligens kan antas medföra så har jag även använt detta arbete till att diskutera utvecklingen inom området och sannolika konsekvenser för exempelvis arbetsmarknad och internationell säkerhetspolitik. Som visades i arbetet finns det mycket som talar för att den mycket snabba utvecklingen inom artificiell intelligens kan antas fortsätta i ett minst lika högt utvecklingstempo framöver. Av patentdatabasen Espacenet framgår att patenten relaterade till begrepp som ”machine learning”, ”neural networks” och ”Deep learning” ökar i mycket snabb takt. Dessutom utvecklas den underliggande datortekniken exponentiellt och det har nyligen gjorts betydande genombrott både när det gäller utvecklingen av mer flexibel och mångsidig artificiell intelligens, exempelvis DeepMinds generella artificiella intelligens, och när det gälla specialiserad hårdvara. Därför menar jag att det sannolikt är rimligt att anta att den generella utvecklingstakten även framöver kommer att vara mycket hög och att det sannolikt för eller senare kommer att utvecklas en artificiell intelligens med kapacitet som överstiger den mänskliga förmågan inom de flesta, eller kanske till och med alla, områden.

Samtidigt kan man också se betydande regionala skillnader inom utvecklingen vid en granskning av patenten inom området. Vid en sökning i patentdatabasen Espacenet konstaterades att USA har flest datarelaterade patent kopplade till begreppen ”machine learning” (46 %) och ”neural networks” (59 %), medan Kina står för 70 % av patenten inom data (IPC: G06) som är kopplade till begreppet ”Deep learning”. Antalet europeiska patent inom området är litet, och i sökningen påträffades endast ett fåtal svenska patent. Om dessa siffror är representativa för utvecklingen i stort indikerar de skillnader i utvecklingen som mycket väl få stora konsekvenser på såväl den globala handeln som den internationella säkerhetspolitiken.

Vidare är det indikationer som tyder på att den snabba utvecklingen inom artificiell intelligens kan få stora samhällskonsekvenser för exempelvis ekonomin och arbetsmarknaderna inom relativt kort tid. Redan idag pågår en massiv utbyggnad i Kina, exempelvis driftsätter företaget Foxconn i snitt ca 10 000 industrirobotar om året. Dessutom har biltillverkaren Tesla meddelat att de avser att producera sin storsatsning Modell 3 i en helautomatiserad fabrik. Som tidigare nämnts finns det forskarrapporter från 2013 och 2014 som indikerar att ungefär hälften av alla arbetstillfällen riskerar att ersättas av robotar och artificiell intelligens inom 10-20 år och efter publiceringen av rapporterna har tekniken utvecklats betydligt.

Den låga mängden europeiska (och inte minst svenska) patent kan därför å ena sidan ses som ett tecken på att omställningen sannolikt kan gå relativt långsamt här, något som skulle kunna tala för en relativt mjuk övergång på arbetsmarknaden. Men å andra sidan kan undersökningen även tolkas som att de europeiska och svenska företagen riskerar att förlora konkurrenskraft när deras internationella konkurrenter går över till allt effektivare och billigare produktion.

Utvecklingen inom artificiell intelligens kan även antas få betydande konsekvenser för de internationella relationerna, exempelvis eftersom autonoma vapensystem kan rubba de säkerhetspolitiska balanserna och skapa osäkerhet kring de ländernas militära kapacitet. Med tanke på de allvarliga konsekvenser som potentiellt kan följa om övergången till en ekonomi baserad på artificiell intelligens görs på ett vårdslöst sätt menar jag att de riktlinjer som The Future of Life Institute tar upp i sitt upprop kan ses som en miniminivå för de hänsyn och övervägningar som behöver göras (Future of life, 2016, Russell, Dewey & Tegmark, 2016).

Avslutande ord

Ett genomgående tema i detta arbete är den mycket snabba utvecklingen inom artificiell intelligens, som kan innebära både mycket stora möjligheter och betydande risker för såväl individer som samhällen och för världssamfundet i stort. Men med hänsyn till både utrymme och sammanhang har många intressanta och viktiga diskussioner om utvecklingen inom artificiell intelligens utlämnats och läsaren rekommenderas därför varmt att ta del av den rapport inom området som skrevs på initiativ av USAs tidigare president Barack Obama:

> https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Rapporten behandlar bland annat potentiella konsekvenser av utvecklingen inom artificiell intelligens och författarna kommer med omfattande rekommendationer på strategier på samhällsnivå. I rapporten konstaterar författarna (Furman *et al*., 2016, s.12):

> Economic theory suggests that there must be gains from innovations, or they would not be adopted. Market forces alone, however, will not ensure that the financial benefits from innovations are broadly shared.

Som jag ser det mycket kloka och tankvärda ord, som jag därför låter avsluta detta arbete.

Källor

Acemoglu D. & Autor D. (2011). *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Handbook of labor economics 4(2011): s.1043-171, http://economics.mit.edu/files/5571 [170130]

Ariley D. (2008). *(O)logiskt – varför smarta människor fattar irrationella beslut*, Ica Bokförlag, Forma Publishing Group AB, Solna

Armstrong S. & Sotala K. (2015). *How We're Predicting AI—or Failing To*, Beyond Artificial Intelligence: The Disappearing Human-Machine Divide (Ed. Romportl J., Kelemen J. & Zackova E.), 2015, s.11-29. Springer International Press, Schweiz.

Arntz M., Gregory T., & Zierahn U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis*, OECD Social, Employment and Migration Working Papers No. 189, http://www.oecd-ilibrary.org/docserver/download/5jlz9h56dvq7-en.pdf?expires=1480994298&id=id&accname=guest&checksum=6DC4B241A91EE860DC391585FF43C51C [170130]

Atkinson R.D. (2015). *The 2015 ITIF Luddite Award Nominees: The Worst of the Year's Worst Innovation Killers*, ITIF, http://www2.itif.org/2015-itif-luddite-award.pdf?_ga=1.109047563.826258778.1483736154 [161210]

Autor D.H. (2015). *Why Are There Still So Many Jobs? The History and Future of Workplace Automation*, Journal of Economic Perspectives, 29(3), 3-30.

Bach C.F. (2015). *International cooperation must drive autonomous cars*, UNECE, https://www.unece.org/info/media/blog/previous-blogs/international-cooperation-must-drive-autonomous-cars.html [161010]

Bitcoin.se (2013). *Vad är en ASIC?*, http://www.bitcoin.se/2013/02/04/vad-ar-en-asic/ [170131]

Bostrom N. (2014). *Superintelligence – Paths, Dangers, Strategies*, Oxford University Press, Oxford, UK.

Bostrom N. & Yudkowsky E. (2014). *The Ethics of Artificial Intelligence*, The Cambridge Handbook of Artificial Intelligence (Ed. Frankish K. & Ramsey W.M.), 2014, s.316-335, Cambridge University Press, Cambridge, UK.

Buchanan M. (2015). *Physics in finance: Trading at the speed of light*, Nature, 518(7538), http://www.nature.com/news/physics-in-finance-trading-at-the-speed-of-light-1.16872 [161216]

Chamary J.V. (2016). *Drones Can Defeat Humans Using Artificial Intelligence*, Forbes 2016-06-28, http://www.forbes.com/sites/jvchamary/2016/06/28/ai-drone/#52ec9e7d6297 [160706]

Clark J. (2016). *Google Cuts Its Giant Electricity Bill With DeepMind-Powered AI*, Bloomberg, 2016-07-19, https://www.bloomberg.com/news/articles/2016-07-19/google-cuts-its-giant-electricity-bill-with-deepmind-powered-ai [161013]

Clemen R.T. & Reilly T. (2001). *Making Hard Decisions*, South-Western Cengage Learning, Mason, Ohio, USA

Cohen R. (2016). *Machine of the human mind*, SVT, http://urplay.se/program/196622-skaparen-av-manskliga-robotar [161010]

Cox D. (2016). *Welcome speech, Harvard Brain + Machines Symposium 1/22/16*, https://www.youtube.com/watch?v=SaMiYa9YnCw&index=2&list=PLfjZYvoyxDtZjm9wJVxQIOU8SvoUBG3sQ [160713]

Cox D. (2015a). *Hippocampus*, MCB80x Fundamentals of Neuroscience, Harvard University, https://www.mcb80x.org/course/the_brain/subcortical_structures/hippocampus [170105]

Cox D. (2015b). *Cerebellum*, MCB80x Fundamentals of Neuroscience, Harvard University, https://www.mcb80x.org/course/the_brain/subcortical_structures/cerebellum [170105]

Cruickshank M. (2015). *China begins construction of all-robot factory*, The Manufacturer, 20150507, http://www.themanufacturer.com/articles/china-begins-construction-of-all-robot-factory/ [170131]

Darwin C. (1861). *On the origin of species*, http://darwin-online.org.uk/converted/pdf/1861_OriginNY_F382.pdf [170105]

Descartes R. (1637). *Discours de la method*, http://classiques.uqac.ca/classiques/Descartes/discours_methode/Discours_methode.pdf [160710]

Deng L. & Yu D. (2014). *Deep Learning: Methods and Applications*, Foundations and Trends in Signal Processing, 7(3–4), 197–387.

Domingos P. (2015). *The Master Algorithm - How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York, USA

Eagleman D. (2015). *The brain with David Eagleman*, Part 4 of 6, Utbildningsradion, http://urplay.se/program/197087-var-manskliga-hjarna-hur-fattar-vi-beslut [161121]

Eklund K. (1995). *Vår ekonomi – En introduktion till samhällsekonomin*, Femte upplagan, Tidens förlag, Stockholm, Sverige

Engel T.A., Chaisangmongkon W., Freedman D.J. & Wang X-J. (2015). *Choice-correlated activity fluctuations underlie learning of neuronal category representation*, Nature Communications, 6(6454), http://www.nature.com/ncomms/2015/150311/ncomms7454/full/ncomms7454.html [160708]

EPO (2017). Espacenet – Patent search, https://worldwide.espacenet.com/ [20170220]

Ferrucci D., Brown E., Chu-Carroll J., Fan J., Gondek D., Kalyanpur A.A., Lally A., Murdock J.W., Nyberg E., Prager J., Schlaefer N., & Welty C. (2010). *The AI Behind Watson — The Technical Article*, AI Magazine, Fall 2010, http://www.aaai.org/Magazine/Watson/watson.php [161217]

Feuillet L., Dufour H. & Pelletier J. (2007). *Brain of a white-collar worker*, The Lancet, 370(9583), 262, http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(07)61127-1/fulltext [170111]

Fischetti M. (2011). *IBM Simulates 4.5 percent of the Human Brain, and All of the Cat Brain*, Scientific American, 2011-10-25, http://www.scientificamerican.com/article/graphic-science-ibm-simulates-4-percent-human-brain-all-of-cat-brain/ [160713]

Forward S. (2008). *Driving Violations - Investigating Forms of Irrational Rationality*, Uppsala universitet, Doktorsavhandling, Institutionen för psykologi, http://uu.diva-portal.org/smash/get/diva2:172720/FULLTEXT01.pdf [160706]

Frey C. & Osborne M. (2013). *The Future of Employment: How Susceptible are Jobs to Computerization*, Oxford University, http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf [160915]

Fung B. (2016). *Elon Musk: Tesla's Model 3 factory could look like an alien warship*, The Washington Post, 160804, https://www.washingtonpost.com/news/the-switch/wp/2016/08/04/the-future-of-car-production-will-be-devoid-of-people-according-to-tesla/?utm_term=.8aaf335e8359 [170131]

Furman J., Holdren J.P., Muñoz C., Smith M. & Zients J. (2016). *Artificial Intelligence, Automation, and the Economy*, Executive office of the President https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF [170120]

Future of life (2016). *Research Priorities for Robust and Beneficial Artificial Intelligence*, http://futureoflife.org/ai-open-letter/ [170105]

Fölster S. (2014). *Vartannat jobb automatiseras inom 20 år – utmaningar för Sverige, Stiftelsen för Strategisk Forskning*, http://stratresearch.se/wp-content/uploads/varannat-jobb-automatiseras.pdf [160420]

Gibney E. (2015). *Inside DeepMind*, Nature Video, 150224 https://www.youtube.com/watch?v=xN1d3qHMIEQ&feature=youtu.be [161012]

Ginsburg S. & Jablonka E. (2007). *The Transition to Experiencing: II. The Evolution of Associative Learning Based on Feelings*, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.183.6158

Grossman L. (2011). *2045: The Year Man Becomes Immortal*, Time Magazine, 2011-02-10, http://content.time.com/time/magazine/article/0,9171,2048299,00.html [161114]

Hassabis D. (2015). *Public Lecture with Google DeepMind's Demis Hassabis*, Royal Television Society, https://www.youtube.com/watch?v=0X-NdPtFKq0 [161013]

Hassabis D. (2016a). *Demis Hassabis: Towards General Artificial Intelligence*, Center for Brains, Minds and Machines (CBMM), https://www.youtube.com/watch?v=vQXAsdMa_8A, [161013]

Hassabis D. (2016b). *Demis Hassabis: Artificial Intelligence and the Future*, Royal Society of Arts, https://www.youtube.com/watch?v=cEL4iR-d4L8 [16101]

Haugeland J. (1985). *Artificial Intelligence: The Very Idea*, The MIT Press, Cambridge Massachusetts, USA, http://lolita.unice.fr/~scheer/cogsci/Haugeland%2089%20-%20Artificial%20intelligence.pdf [161025]

HealthTap (2014). *HealthTap: Helping Millions Feel Good*, https://www.youtube.com/watch?v=euXY8YJSL84 [170127]

Heyns C. (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, Christof Heyns, A/HRC/23/47, 130409, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf [160410]

Hofstadter D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic books, USA.

Google (2016). https://www.google.com/selfdrivingcar/ [160919]

IAI (2013). *HAROP*, http://www.iai.co.il/2013/36694-46079-EN/Business_Areas_Land.aspx [161128]

Intel (2016). *Fueling Innovation We Love and Depend On*, http://www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html [160701]

ITIF (2016a). *Artificial Intelligence Alarmists Win ITIF's Annual Luddite Award*, https://itif.org/publications/2016/01/19/artificial-intelligence-alarmists-win-itif%E2%80%99s-annual-luddite-award[170107]

ITIF (2016b). *About ITIF: A Champion for Innovation*, https://itif.org/about [170107]

Janlert L.E. (2015). *Tänkande och beräkning – en inledning till datavetenskap och kognitionsvetenskap*, Studentlitteratur AB, Lund

Jaumotte F. & Tytell I. (2007). *How Has The Globalization of Labor Affected the Labor Income Share in Advanced Countries?*, IMF, https://www.imf.org/external/pubs/ft/wp/2007/wp07298.pdf [170115]

Kahneman D. (2013). *Tänka, snabbt och långsamt*, Volante, Stockholm

Kiat O.K. (2016). *The emergence of Artificial Intelligence in the stock market*, e27, 160614, https://e27.co/the-emergence-of-artificial-intelligence-in-trading-20160614/ [161216]

Kurzweil R. (2005). *The Singularity Is Near – When humans transcend biology*, Viking Penguin, New York, USA

Kurzweil R. (2016). *Ray Kurzweil - The Future of Technology*, https://www.youtube.com/watch?v=9yCW_9NIAUw [170129]

Kuz (2016). *DeepMind-Atari-Deep-Q-Learner*, https://github.com/kuz/DeepMind-Atari-Deep-Q-Learner [161130]

Lake B.M., Ullman T.D., Tenenbaum J.B. & Gershman S.J. (2016). *Building Machines That Learn and Think Like People*, http://arxiv.org/abs/1604.00289 [161227]

Larsson L. (2016). *Världen tar stormsteg mot tänkande maskiner*, Dagens Nyheter 161209, http://www.dn.se/ekonomi/varlden-tar-stormsteg-mot-tankande-maskiner/ [161217]

Lee C.H. & Hwang A. (2016). *Foxconn boosting automated production in China*, Digitimes, 161230, http://www.digitimes.com/news/a20161229PD206.html [170110]

Manyika J., Chui M., Bughin J., Dobbs R., Bisson P. & Marrs A. (2013). *Disruptive Technologies: Advances that will transform life, business, and the global economy*, McKinsey Global Institute, http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies [170105]

Mearsheimer J.J. (2001). *The tragedy of great power politics*, W W Norton & Company, New York and London.

Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., Graves A., Riedmiller M., Fidjeland A.K., Ostrovski G., Petersen S., Beattie C., Amir Sadik A., Antonoglou I., King H., Kumaran D., Wierstra D., Legg S. & Hassabis D. (2015). *Human-level control through deep reinforcement learning*, Nature, 518(7540), 529–533, http://www.nature.com/nature/journal/v518/n7540/pdf/nature14236.pdf [160912]

Mordvintsev A., Olah C. & Tyka M. (2015). *Inceptionism: Going deeper into Neural Networks*, Google Research Blog, https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html [170213]

Mowle T.S. (2003). *Worldviews in Foreign Policy: Realism, Liberalism, and External Conflict*, Political Psychology, 24(3), 561-592.

Munz M., Gobert D., Schohl A., Poquerusse J., Podgorski K., Spratt P., Ruthazer E.S., (2014). *Rapid hebbian axonal remodeling mediated by visual stimulation*. Science, 344 (6186), 904, https://www.sciencedaily.com/releases/2014/05/140528104953.htm [160807]

Mölne V. (2016). *Ersätter 60 000 anställda med robotar*, Dagens industri, 160526, http://www.di.se/artiklar/2016/5/26/ersatter-60000-anstallda-med-robotar/ [160526]

Nicolelis M. (2011). *Beyond Boundaries: The New Neuroscience of Connecting Brains with Machines - And How It Will Change Our Lives*, Times Books, Henry Holt and Company, LLC, Newe York, USA

Nilsson N.J. (2010). *The Quest for Artificial Intelligence*, Cambridge University Press, New York, USA

Nilzén R. (2008). *Teorin om spegelneuron förklarar förmågan till empati*, Läkartidningen, 2008(32-33), http://ww2.lakartidningen.se/store/articlepdf/1/10018/LKT0832s2193_2195.pdf [160706]

Noyes K. (2016). *Så har IBM tagit Watson från Jeopardy till big business*, Computer Sweden, 2016-10-02, http://computersweden.idg.se/2.2683/1.666504/ibm-watson-utveckling [161006]

Olsson Jeffery M. (2016). *Martin Ford: "Det mesta har ett slut"*, Dagens Industri 160920, http://www.di.se/artiklar/2016/9/20/martin-ford-det-mesta-har-ett-slut/ [160921]

Osborne J. (2016). *Google's Tensor Processing Unit explained: this is what the future of computing looks like*, TechRadar, 160822, http://www.techradar.com/news/computing-components/processors/google-s-tensor-processing-unit-explained-this-is-what-the-future-of-computing-looks-like-1326915 [160914]

Parkin S. (2015). *Killer robots: The soliders that never sleep*, BBC – Future, 150716, http://www.bbc.com/future/story/20150715-killer-robots-the-soldiers-that-never-sleep [161217]

Pitt J. (2010). *General Intelligence via Cognitive Synergy*, OpenCog, http://opencog.org/theory/ [161013]

Porter M.E. (1979). *How competitive forces shape strategy*, Harvard Business Review, March-April.

Pratt G.A. (2015). *Is a Cambrian Explosion Coming for Robotics?*, The Journal of Economic Perspectives, 29(3), 51-60, http://www.jstor.org/stable/43550120?seq=1#page_scan_tab_contents [170201]

Prigg M. (2014). *Who goes there? Samsung unveils robot sentry that can kill from two miles away*, Daily Mail, 140915, http://www.dailymail.co.uk/sciencetech/article-2756847/Who-goes-Samsung-reveals-robot-sentry-set-eye-North-Korea.html [161128]

Purves D., Augustine G.J., Fitzpatrick D., Hall C.W., LaMantia A-S. & White L.E., (2012). *Neuroscience*, Fifth edition, Sinauer Associates Inc, Sunderland, Massachusetts, USA

Purves D., Cabeza R., Huettel S.A., LaBar K.S., Platt M.L. & Woldorff M.G. (2013). *Principles of Cognitive Neuroscience*, Second edition, Sinauer Associates Inc, Sunderland, Massachusetts, USA

Realsentrygun.com (2014). *Realsentrygun.com – the most experienced name in sentry gun systems*, http://www.realsentrygun.com/ [170127]

Riel T. (1993). *Animate characters by evolving them*, https://www.ted.com/talks/torsten_reil_studies_biology_to_make_animation#t-199004 [161115]

Roos U. (2010). *The aging society*, BIL Sweden, Presentation på MHFs konferens Tylösandsseminariet 2010, http://www.mhf.se/client/files//content/projekt_vi_driver/Tylosandseminariet/presentationer/13.40_Ulf_Roos.pdf [170125]

Rose C. (2016). *Artificial intelligence*, CBS 60 Minutes, http://www.cbsnews.com/videos/artificial-intelligence/ [161012]

Russell S. (2015). Intervju publicerad på tidningen Natures hemsida, ljudupptagning, http://www.nature.com/news/robotics-ethics-of-artificial-intelligence-1.17611 [161128]

Russell S., Dewey D. & Tegmark M. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence*, AI Magazine, Winter 2015, http://futureoflife.org/data/documents/research_priorities.pdf, [160412]

Russell S. & Norvig P. (2010). *Artificial Intelligence – A Modern Approach*, Third edition, Pearson Education Inc., Upper Saddle River, New Jersey, USA

Salmon F. & Stokes J. (2010). *Algorithms take control of Wall Street*, Wired Magazine, 101227, https://www.wired.com/2010/12/ff_ai_flashtrading/ [161217]

Senitent (2016). *Sentient Investment Management*, http://www.sentient.ai/sentient-investment-management/ [161216]

Socialstyrelsen (2008). *Vårdskador inom somatisk slutenvård*, https://www.skane.se/upload/Webbplatser/Utvecklingscentrum/dokument/Socialstyrelsen%20v%C3%A5rdskadem%C3%A4tning.pdf [160706]

Stanton G.H. (2013). *The Ten Stages of Genocide*, Genocide Watch, http://genocidewatch.org/images/Ten_Stages_of_Genocide_by_Gregory_Stanton.pdf [170202]

Steadman I. (2013). *IBM's Watson is better at diagnosing cancer than human doctors*, Wired 2013-02-11, http://www.wired.co.uk/article/ibm-watson-medical-doctor [160706]

Ström V. (2016). *Fyra nya robotfirmor startar i Sverige*. Dagen Industri, 160708, http://digital.di.se/artikel/fyra-nya-robotfirmor-startar-i-sverige [160709]

Sveriges Riksbank (2015). *Den svenska finansmarknaden*, Stockholm, Sverige

Söderström R. (2011). *Algoritmisk handel med aktier: några rättsliga implikationer*, Tidsskrift for Selskabsret, 2011(3), http://www.diva-portal.org/smash/get/diva2:482691/FULLTEXT02 [160921]

Tesla (2016). *All Tesla Cars Being Produced Now Have Full Self-Driving Hardware*, 161019, https://www.tesla.com/sv_SE/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware?redirect=no [161216]

Transportstyrelsen (2016). *Olycksstatistik – nationell statistik* http://www.transportstyrelsen.se/sv/vagtrafik/statistik-och-register/Vag/Olycksstatistik/Polisrapporterad-statistik/Nationell-statistik/ [160706]

Turing A.M. (1950). *Computing Machinery and Intelligence*, Mind 49, 433-460

Tversky A. & Kahneman D. (1974). *Judgment under Uncertainty: Heuristics and Biases*, Science, 185 (4157), 1124-1131.

Törnros R. (2015). *Vi kör Tesla Model S P90D med autopilot*, Teknikens Värld, http://teknikensvarld.se/vi-kor-tesla-model-s-p90d-med-autopilot-214640/ [160919]

United States Air Force (2009). *Unmanned Aircraft Systems Flight Plan 2009-2047*, https://fas.org/irp/program/collect/uas_2009.pdf [160706]

United States Department of Defense (2013). *Unmanned Systems Integrated Roadmap FY2013-2038*, http://www.defense.gov/Portals/1/Documents/pubs/DOD-USRM-2013.pdf [160423]

Verscend (2017). *Optimizing insights and data for smarter solutions*, https://www.verscend.com/solutions [170124]

Vinge V. (1993). *The Coming Technological Singularity: How to Survive in the Post-Human Era*, http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html [161115]

Volkswagen Group Sverige AB (2016). *Audi fortsätter tester med självkörande "Jack" på Autobahn*, Pressmeddelande 160519, http://www.volkswagengroup.se/sv/Mediarummet/Pressrelease---visningsida/?pid=1150713 [161120]

Vonko D. (2016). *Neural Networks: Forecasting Profits*, http://www.investopedia.com/articles/trading/06/neuralnetworks.asp [161015]

Wemmenhag H. (2016). *Scania visar upp självkörande lastbilar*, Dagens Industri, 160528, http://www.di.se/artiklar/2016/5/28/scania-visar-upp-sjalvkorande-lastbilar/ [161015]

Weyand T., Kostrikov, I. & Philbin, J. (2016). *PlaNet - Photo Geolocation with Convolutional Neural Networks*, European Conference on Computer Vision (ECCV), https://arxiv.org/pdf/1602.05314v1.pdf [161026]

WEF (2017a). *Davos 2017 - An Insight, An Idea with Sergey Brin*, https://www.youtube.com/watch?v=ffvu6Mr1SVc [170127]

Wu J.R. (2017). *Foxconn CEO says investment for display plant in U.S. would exceed $7 billion*, Reuters, 170122, http://www.reuters.com/article/us-taiwan-foxconn-idUSKBN1560JP [170131]

WWF (2016). *Rekordtidig Overshoot Day - Jordens resurser tar slut 8 augusti*, 160808, http://www.wwf.se/press/aktuellt/1658512-rekordtidig-overshoot-day-jordens-resurser-tar-slut-8-augusti [161130]

Yusof M.H.M. & Mokhtar M.R. (2016). *A Review of Predictive Analytic Applications of Bayesian Network*, International Journal on Advanced Science, Engineering and Information Technology. 6(6), 857-867 http://ijaseit.insightsociety.org/index.php?option=com_content&view=article&id=9&Itemid=1&article_id=1382 [170120]

Yuste R. (2015). *Rafael Yuste: Exploring brain architecture and function with optical methods*, Yustelab, University of Columbia, https://www.youtube.com/watch?v=cw0p1KHbvQA [160915]