

Algoritmisk etik

Vad policydokument berättar om
synen på artificiell intelligens



LUNDS
UNIVERSITET

Examensarbete | Digitala Kulturer | VT-2019

Författare Björn K Fritz | Handledare Robert Willim

Humanistiska och teologiska fakulteterna | Institutionen för kulturvetenskaper

Abstract

The aim of this thesis is to look at the attitude among governments and large corporations towards algorithms in general and with a special focus on artificial intelligens. This is done by examining five policy documents produced by three governmental branches and two private companies, all of them from Europe and the US. The assumption is that policy is one of the methods that those in power can use to form norms and behavior that is in line with their agenda, therefore an examination of these policy texts could reveal the obstacles they have identified toward their goals. Depending on how these obstacles, or problems, are represented in a discourse as algorithms, a policy offers a solution. By critically analysing the problem representation instead of the solution the view of the discourse, as seen by the policy maker, is questioned and some premise that supports the problem view might fall. In doing so the representation of the problem may turn out to be different, along with the solution.

The result of the analysis sugest that the governmental branches mainly discuss algorithms and artificial intelligence in terms of economy, the privacy and integrity of the people is also important as long as it does not interfere with market growth. The policys seem to rest on two premisses, the belief in artificial intelligence as a disruptive force and the promise of technology as a solution of many of mankind's problems.

Keywords: Algorithms, Artificial intelligence, Public policy, Ethics, Future society, Inequality, Bias.

1 - Inledning	1
1.1 Bakgrund	1
1.2 Syfte	2
1.3 Avgränsningar	2
1.4 Tidigare forskning	2
1.5 Teorier - begreppsdiskussion	3
1.5.1 Diskurs	3
1.5.2 Policy	4
1.5.3 Etik	5
1.6 Metodval	6
1.6.1 Policydokument som studieobjekt	6
1.6.2 Inspiration för metod och materialval	7
1.7 Forskningsetik	8
2 - Algoritmer och artificiell intelligens	9
2.1 Artificiell (Generell) Intelligens	9
2.2 Historisk tillbakablick	9
2.3 Litteraturens bidrag	10
2.4 Artificiell intelligens - hot eller möjlighet?	11
2.5 Algoritmer skapar förutsättningar för en ny ekonomi	12
3 - Sammanställning av policydokumenten	14
3.1 Offentliga aktörer	14
3.1.1 Europeiska kommissionen	14
3.1.2 Executive Office of the President (of the United States)	14
3.1.3 Government office for science (Storbritannien)	15
3.2 Privata aktörer	16
3.2.1 IBM Cognitive Principles	16
3.2.2 Artificial Intelligence; The Public Policy Opportunity (Intel)	17
3.3 Sammanställning av dokumenten	18
4 - Analys	19
4.1 Vilka problem representeras?	19
4.2 Regler och transparens krävs för att artificiell intelligens skall nå sin fulla potential	20
4.3 Automatisering och artificiell intelligens är en kapprustning med allt högre insatser	22
4.4 Satsning på AI skapar nya kreativa yrkesroller istället för de som förloras.	25
4.5 Algoritmer, för vem?	27
5 - Slutsatser	29
5.1 Avslutande reflektion	30
Källhänvisningar	32
Policydokument	33

1 - Inledning

Artificiell intelligens har gått från att vara ett begrepp som främst kittlat fantasin hos läsare av science fiction-litteratur eller biobesökare på filmer som rymdäventyret *2001* och *iRobot* till att idag diskuteras i anslutning till de flesta samhällsfrågor. Artificiell intelligens eller förkortningen AI används för att sälja tvättmaskiner såväl som ideologiska budskap, självkörande intelligenta bilar säljs in som en möjlighet att minska både trafikolyckor och utsläpp samtidigt som de problematiseras utifrån etik och personlig integritet. Att se sin brödrost som intelligent kan även väcka existensiella problem, ett bättre alternativ är att säga att den är styrd av algoritmer. Titeln på detta uppsatsarbete, *Algoritmisk etik*, är valt utifrån den övergripande karaktär begreppet algoritmers har, men då artificiell intelligens idag ofta används som synonymt med algoritmer kommer även denna uppsats att reflektera detta.

Liksom industrialismen med massproduktion och effektivare infrastruktur kan beskrivas som en möjlighet att erbjuda alla samma möjligheter till ett materiellt bra liv och möjlighet till självförverkligande, kan även algoritmer genom digitaliseringen och Internets framväxt framstå som möjligheternas teknik. En jämställd plattform där alla kan göra sin röst hörd, en demokratisk mötesplats där informationen görs tillgänglig för alla, på lika villkor. Historien har dock visat att tekniken främst gynnar dem med kunskap om och tillgång till den, att teknik möjliggör men även begränsar. När nu algoritmer framställs som lösningen på många av dagens problemområden, demokrati, klimat och hälsa, måste frågan ställas - för vem löser de problemen och vilka betalar priset?

1.1 Bakgrund

Kultur kan ses som beteenden och den samlade kunskapen hos en grupp människor, kunskap och färdigheter vilka de värdesätter och därför förmedlar till de nytillkomna i gruppen (Schein i Strannegård och Eriksson-Zetterquist 2011:45). Att se organisationer som kulturer ger därigenom möjligheten att studera dem med hjälp av hermeneutiska verktyg som text- och diskursanalys. Policydokument är ett sätt för organisationer att förmedla den gemensamma värdegrunden och målen de strävar mot, deras policy i frågor kring artificiell intelligens bör då också gälla hela organisationen och inte bara de vilka är involverade i teknisk utveckling.

Denna uppsats kommer att baseras på en analys av policydokument, och speglar den växande betydelse artificiell intelligens förväntas få i samhället, då denna form av uppsats har begränsat utrymme finns också behov av avgränsningar. Att välja ut dokument är i sig självt något som påverkar resultatet och måste därför både göras med omsorg och sedan finnas med i analysarbetet som en parameter. Trots ämnets aktualitet är det svårt att finna publicerade dokument hos de stora internetföretagen. Valet föll slutligen på två privata företag och tre offentliga organisationer. IBM och Intel är utvalda att representera industrin, EU-kommissionen och stadsapparaterna i Storbritannien och USA det offentliga.

1.2 Syfte

Syftet med denna uppsats är att visa hur synen på algoritmer, framför allt i form av begreppet artificiell intelligens, hos offentliga och privata organisationer kan belysas utifrån policydokument, samt i vilken utsträckning diskursens bakomliggande premisserna motsvarar de problem som uttrycks. Uppsatsen tar stöd i den relevanta forskning, framför allt den inom humaniora, vilken studerar artificiell intelligens och dess påverkan på människors livsvillkor.

- Vad kan policydokument om artificiell intelligens berätta om organisationernas syn på frågor som personlig integritet, etik och moral?
- Vilken effekt kan dessa policydokument förväntas ha på samhället?

1.3 Avgränsningar

De utvalda dokumenten är samtliga från stora organisationer och företag, denna avgränsning är gjord utifrån antagandet att de då också representerar organisationer vilka bör ha ett reellt inflytande på utvecklingen av både tekniken och användarna.

Att utgå från institutioner och företag vilka i viss mån kontrollerar utvecklingens riktning påverkar även fokuset på de frågor som behandlas, akademiska och filosofiska frågor som identitetsskapande och kulturell påverkan är inte klart utskrivna i dokumenten. De reflekterar istället behovet från beslutsfattarnas att förmedla sin syn på att nå budgetbalans och internationell konkurrenskraft, hur organisationernas övergripande mål uppnås. Det är därför viktigt att även uppmärksamma de frågeställningar vilka utelämnas i dokumenten, eller endast berörs i förbigående. Dessa kan visa på hur rådande strukturer förhindrar nya tankar, och om de folkvalda politikerna bortser från experters råd och istället tar populära beslut i förhoppning om att bli omvalda (Cairney 2012:67). Samtliga organisationer har också sitt ursprung framför allt i en västerländsk kontext, vilket gör att resultatet kommer att färgas utifrån dessa länders värderingar. Kina är en av de länder som satsar mest på forskning och utveckling av algoritmer, men att finna policydokument i ämnet från Kinesiska organisationer har visat sig vara svårt. Samtidigt gör denna avgränsning att just ett västerländskt perspektiv på algoritmer bör framträda desto tydligare.

1.4 Tidigare forskning

Artificiell intelligens som teknik kan uppfattas som neutral, att det producerade resultat bygger på logiska och objektiva processer. Detta är något som bland andra Safiya Noble starkt ifrågasätter, i sin forskning på Googles sökmotor visar hon hur algoritmerna många gånger bygger på fördomar och därmed uppvisar det som benämns bias, vilken är både rasistisk och sexistisk. Detta sker enligt henne på

grund av Googles roll–inte som uppslagsverk–utan som vinstdrivande företag, informationen de presenterar måste vara relevant–inte nödvändigtvis riktig–för den som söker (Noble 2018:149,152). Denna bias behöver inte vara oavsiktlig, vilket Virginia Eubanks påminner oss om i boken *Automating Inequality*. I den jämför hon många av dagens sociala välfärdssystem med 1800-talets fattighus i USA, där de boende tvingades avsvära sig sina rättigheter och sedan blev utnyttjade av ägarna (Eubanks. 2017:18). Samhället undgick att ta de ekonomiska kostnaderna som följer av det sociala ansvaret för de sämre lottade och lade sig därför inte i verksamheten vilken på sina ställen fortsatte långt in på 1900-talet. På samma sätt ser hon hur de system idag, vilka ska agera sociala skyddsnät, kräver att den som söker stöd också låter sig skärskådas utan begränsningar, och så länge inte systemen uttryckligen är konstruerade för att motverka bias i samhället kommer de istället att förstärka dem (ibid 190). Eubanks lanserar i sin bok begreppet *digitalt fattighus* som beskrivning av situationen för stora delar av den lägre arbetarklassen, genom bias begränsas jämlikhet både i form av frihet och möjligheter (ibid 194). Även svenska forskare intresserar sig för hur etiska värden byggs in i algoritmerna, Olof Sundin menar i en intervju att även om en artificiell intelligens gör allting rätt så är "[...]principerna för beräkningen en mänsklig konstruktion." Sundin antyder här att en rent objektiv utgångspunkt blir svår att finna, istället förespråkar han "transparens och algoritmisk ansvarsskyldighet", vilket möjliggör både förståelse för och möjlighet att bestrida ett resultat (Andersson Schwarz, Larsson 2019:105).

1.5 Teorier - begreppsdiskussion

En del av de teorier som förekommer i uppsatsen kommer utifrån tidigare forskning, den kritiska ingången till bias hos algoritmer vilket bland andra Noble och Eubanks representerar. I en diskussion kring policydokument finns också frågan om makt, vilka utövar den och vilka berörs. Policy som begreppet saknas i det svenska språket och går inte direkt att översätta men dess grund i politiken finns där. Slutligen är etiken viktig att diskutera i detta sammanhang, går det att skapa en utveckling inom artificiell intelligens vilken tar hänsyn till de etiska och moraliska värden som format den moderna människan?

1.5.1 Diskurs

Med en kritisk diskursanalys om algoritmer och dess påverkan på kulturella normer är det naturligt att hänvisa till Michel Foucault och hans diskursteorier om maktförhållanden och kunskap. I det primära arbetet läggs tyngdpunkten av analysarbetet på språkets användning i diskussionerna kring algoritmer (Boréus 2015:177). Med utgångspunkten i en diskursanalys avgränsas även studieobjektet, algoritmer, i tid och rum, vilket även begränsar hur vi kan beskriva och förklara dess återverkningar (Börjesson, 2003:34). En viktig del av en diskurs är enligt Foucault hur makt fördelas, någons frihet blir någon annans begränsning vilket inte behöver vara en medveten handling. Istället kan dessa

utestängningsmekanismer grundas på vad som anses förbjudet, sjukt eller tradition inom en diskurs vilket i sin tur är baserat på det vilket anses vara försanthållet. (Bergström, Boréus 2005:305-314).

Foucault grundar sin teori om hur maktförhållanden och normer i samhället vidmakthålls utifrån funktionen i Jeremy Bentham's skapelse, panopticon (cirkulärt fängelse), där ansvaret för att upprätthålla maktbalansen hamnar hos den betraktade fången. I en analogin med dagens digitala plattformar får algoritmerna representera fängvaktarna som en del av övervakningssamhället, och frågan är hur detta kan påverka användares beteende (Foucault 1991:200).

Vad representerar då en problematisering av övervakning, vad är problemet? Foucault ansåg att en problematisering inte representerar något objekt utan endast är en idé vilken kan utsättas för till exempel etiska, vetenskapliga och politiska frågeställningar. Det går inte att ifrågasätta det etiska med en kamera eller en GPS-kretsen i mobilen, det är när vi kombinerar dess användning med våra dagliga rutiner som diskursen övervakningssamhället skapas. Det är först då vi kan diskutera rätt och fel i frågor om personlig integritet och en trygg boendemiljö (Bacchi 2012:4).

1.5.2 Policy

Vilka är då dessa normer vilka Foucault menar styr handlandet och vem bestämmer om dessa? Policy är ett sätt för de som innehar maktpositioner att skilja det normala från det udda, ett annat kan till exempel vara politiska ställningstaganden. En avgörande skillnad mellan dessa är att politik kan uppfattas som oeniga krafter vilka kämpar om majoritetens gillande för sin ståndpunkt medan policybeslut snarare framstår som väl underbyggda beslut vilka redan är fattade (Colebatch. 2009:20). Oavsett vilken av dessa metoder som används styrs de normativa värderingarna utifrån en maktposition, möjligheten och rätten att utöva denna makt ges i form av ett samtycke från de som makten utövas på. Detta leder i en demokrati till att de som ger sitt samtycke gör detta utifrån normer vilka skapats genom policybeslut och andra politiska styrmedel som de valda makthavarna skapar.

Policy som instrument för styrning bygger på tre delar: regelverk, auktoritet och expertis. Genom att ställa upp klara regler varken kan eller behöver enskilda medarbetare eller avdelningar ta egna beslut i frågor där dessa omfattas av en officiell policy. Den uppställda policyn förväntas vara framarbetad av en grupp med specialkompetens inom det område som omfattas, dess legitimitet förstärks sedan utifrån att vara sanktionerad från toppen av organisationens hierarkin. Dessa tre delar har inte nödvändigtvis jämlig påverkan på den slutgiltiga policyn, till exempel kan en stark auktoritet bortse från expertisens råd och istället basera en policy på tidigare normer eller privata intressen (Colebatch 2009:8). Att använda policy i beslutsprocessen är inte heller oproblematiskt, det är inte ovanligt att mer än en policy kan beröra samma fråga. För en arbetsgivare som undersöker möjligheten att effektivisera verksamheten kan det vara nödvändigt att ta hänsyn till såväl arbetsmiljö- klimat- som jämställdhetspolicy. Policy kan även ses som ett sätt för de styrande, beslutsfattarna, att utesluta meningsmotståndare från att delta i

beslutsprocessen. Genom att på EU- eller federal nivå definiera vissa idéer som viktiga och utesluta andra från diskussionen kan de forma utvecklingen av en diskurs som AI-teknik utan att behöva ta beslut i enskilda fall (Colebatch 2009:33, Cairney 2012:63). Detta leder i viss mån tillbaka till Foucaults teori om kontroll genom panoptikon, de anpassar sitt beteende efter rådande policy och kommer även att kontrollera att andra följer den (Cairney. 2012:61). Denna syn på rådande policy som en gemensam överenskommelse för allas bästa kan överskugga individens insikt om de egna behov, vilket i förlängningen kan ses som ett sätt att skapa en bias vilken främst gynnar de styrande (ibid 67).

1.5.3 Etik

På områden som vård, skola och omsorg har etiska riktlinjer en given plats, även pressen har sin etiska nämnd liksom börsföretagens förhållande gentemot sina aktieägare inte endast baseras på lagar utan även regleras genom etiska regler. Dessa regler och den moral de bygger på är grundad i en diskurs, att köpa sig fördelar är inom många områden accepterat men att betala för bättre betyg i skolan eller för att gå före i operationsköer kan ses som högst oetiskt.

Moralen kan ses utifrån ett evolutionärt perspektiv som något gynnsamt för människans utveckling. En metastudie utförd av antropologer vid universitetet i Oxford, vilka studerat moral utifrån ett globalt perspektiv, visar på sju universella uppfattningar vilka gynnar mellanmänniskt samarbete. Genom historien har hantering av resurser som att (1) hjälpa släktingar, (2) lojalitet mot gruppen och (3) gentjänster gynnat den egna gruppens utveckling. När konflikter uppstår hanteras dessa moraliskt korrekt främst genom utmaningar, och avgörs genom att den (4) dominante (modige) tillåts vinna genom att den (5) fredlige ger sig. Slutligen binds mänskliga civilisationer samman av en moralisk uppfattning av (6) rättvisa och (7) rätt till egendom. Dessa kan ses som universala i utsträckningen att även om dessa egenskaper inte alltid ses som moraliskt rätt så ses de aldrig som fel (Scott Curry. Austin Mullins. Whitehouse. 2019). Detta har hjälpt mänskligheten att utvecklas till där vi är idag, dessa bygger samtliga på mellanmännisklig interaktion. Hur påverkas då vår lojalitet eller syn på rättvisa när budskap medieras genom eller till och med emanerar från algoritmer?

Det vi inte förstår, det vilket är allt för tekniskt avancerat, kan uppfattas som något opåverkligt och deterministiskt. Användare av ett sådant system kan komma att fränsäga sig allt moraliskt ansvar för slutresultatet, detta läggs istället på systemutvecklarna vilka ensamma får ansvara för utvecklingens inriktningen. När dessa system blir allt för stora förändras ofta dess syftet från att serva människor till att, genom finlipade tekniska system, effektivisera driften för att först och främst öka den ekonomiska vinning. De inblandade i systemen, både i företaget och dess användare reduceras då lätt till kuglar i maskineriet och ting som skall hanteras. Moraliskt kan detta försvaras genom generalisering — det är bara så man gör — det personliga ansvaret ersätts med tilltron till auktoritetens ansvar (Collste 2002:10-12). Att använda en sökmotor, som Google, är troligen för de flesta jämförbart med att köra bil eller använda en mikrovågsugn, i det att vi främst har ett syfte vi vill uppnå. Varken att hitta den

bästa fiskrestaurangen i Örebro, köra bilen dit eller istället värma fiskpinnar kräver kunskap i hur tekniken fungerar under huven. Vi överlåter dessa detaljer åt tillverkarna, experterna, likaså ansvaret för de brister systemen kan vara behäftade med.

Det är inte alltid tanken som räknas, är det oetiskt av Facebook att med hjälp av algoritmer bedöma användarnas känslomässiga tillstånd även om det kan rädda några från att ta sitt liv? Det finns två huvudtyper av normativa etiska teorier, konsekvensetisk och deontologisk. Skillnaden dem emellan är den bedömningsgrund som används. Vid en konsekvensetisk bedömning betraktas handlingens resultat medan den deontologiska endast baseras på handlingen i sig (Collste 2002:32). Detta aktualiseras med synen på dagens avancerade system som mer eller mindre svarta lådor, vi vet vad vi matar in och får ett resultat, men saknar kunskap om processen emellan. Utan inblick i algoritmernas funktion, vilka faktorer den baserar sin slutledning på, finns bara den konsekvensetiska bedömningen kvar. Frågan är då hur ett etiskt regelverk kan skapas baserat på resultatet, istället för den inom områden som ekonomi och tillverkning brukliga regleringen av processen vilka leder fram till det?

1.6 Metodval

Att undersöka algoritmer är att studera ett fenomen vilket saknar fysisk närvaro men vars effekter i högsta grad är påtagliga i människors vardag.

Metodvalet för denna uppsats följer av syftet att undersöka större organisationers syn dels på artificiell intelligens och dess tillämpningar men även vilka förväntningar dessa beslutsfattare har på artificiell intelligens. Istället för att förstå denna diskurs genom att enbart studera aktuell forskning, analyseras fem organisationers förhållningssätt till artificiell intelligens att genom att kritiskt granska deras policydokument som behandlar artificiell intelligens.

1.6.1 Policydokument som studieobjekt

När denna uppsats skrivs har företag inom social media upprepade gånger uppmärksamats för vårdslös hantering av sina användares personliga data. Det främsta exemplet på detta är den så kallade *Cambridge Analytica-skandalen* där Facebook misstänks ha medverkat till att tiotusentals användare utsatts för otillbörlig politisk påverkan både i samband med presidentvalet i USA 2016 och den brittiska omröstningen om utträde ur EU, *Brexit* (SVT 2018).

Förutom tidigare forskning i form av litteratur, artiklar och annat skrivet material kommer policydokument att vara den huvudsakliga källan till information. Dessa är intressanta att studera utifrån hur de väljer att förmedla sin syn på artificiell intelligens i verksamheten, speciellt hur de formulerar etiska frågeställningar i förhållande till det som idag är känt om hur företag som Facebook använder algoritmer på sätt vilket klart bryter mot både etik och moral. En metod av typen *critical discourse analysis* (CDA) kan med fördel användas för att analysera textdokument. Dokumenten och

uttalanden från företagen kan då analyseras utifrån den språkliga uppbyggnaden istället för sitt innehåll. Förespråkare för CDA hävdar att det utelämnade kan vara lika talande som det vilket faktiskt står där, att underförstådda premisser inte skall tas för givna utan kritiskt granskas (Machin, Mayr 2013:4).

I denna uppsats används framför allt en metodik speciellt utarbetad för att genom analys av policydokument närma sig en diskurs, Carol Lee Bacchi ställer frågan ”Vad är problemet så som det representeras?” (”What’s the problem represented to be” förkortat WPR). Hon menar att det finns en direkt koppling mellan hur vi diskuterar och framställer en diskurs och vad problemet är med den. Metoden går i korthet ut på att undersöka vilka underförstådda premisser som finns, vilka försanthållanden skapas som en följd av problemformuleringen, vad ses som oproblematiskt och hur lösningen förändras om problemet omformuleras (Boréus 2015:181-182).

Eftersom det idag skrivs policydokument vilka behandlar artificiell intelligens visar det på att det finns problem i samband med denna teknik, hur de utformas och vilka aspekter de tar upp beskriver också vilka problemområdena är. Då är det också viktigt att reflektera över de val som görs, vad som är önskvärt och vad som skall motverkas, vilken som är den rätta vägen framåt och var fallgroparna är. Då detta ofta är baserat på normer som finns i samhället och utifrån kunskap experter på området besitter är det också något som förändras med tiden (Colebatch 2009:30). Frågan är vilka normer som skapar den policy vilken behandlar användning och utveckling av dagens algoritmer, är teknikutvecklingen så snabb att policybeslut baseras på föråldrad kunskap? Om denna reglering istället lämnas åt marknaden själv, går det att utnyttja plattformarnas egna ranking och rekommendations-system för att på så sätt lämna över makten till konsumenterna? Detta skapar dock ny problematik i tidigare anda, bristen på transparens, normaliseringen som sker genom vad som går att betygsätta samt bristen på koordinering av samhällsfunktioner som påverkar varandra (Van Dijck, Poell, De Waal 2018:90-95).

Plattformsekonomin behöver regleras liksom områden som elektricitet, transporter och banker för att motverka till exempel snedvriden konkurrens, speciell hänsyn behöver tas till dess snabba utveckling för att inte riskera att de policybeslut som tas försöker reglera en verklighet vilken inte längre existerar.

1.6.2 Inspiration för metod och materialval

För att finna återkommande problemformuleringar används i uppsatsen olika former av policydokument, från offentliga myndigheter såväl som privata företag. Inspirationen till detta sätt att använda Bacchis metod (WPR) är taget från informationsvetaren Fredrik Hanells artikel vilken behandlar problematiseringen av digital kompetens i den svenska lärarutbildningen (2018). I den har Hanell utifrån fyra statliga policydokument identifierat fem gemensamma teman vilka ses som den officiella verklighet som problemställningar och lösningsförslag formuleras utifrån. Dessa utgår från att lärarnas bristande förmåga att förmedla digital kompetens hotar att försvaga Sveriges position som digitalt föregångsland. Hanell flyttar fokus från ekonomiska argument för att visa på betydelsen av

andra delar av utbildningen, för demokratin, samhället och kulturell mognad, områden vilka är nog så viktiga för framtiden.

1.7 Forskningsetik

Det är lämpligt att en uppsats som behandlar etik även genomförs på ett etisk korrekt sätt. Denna etik kan ses utifrån två aspekter, dels skall all forskning bedrivas utifrån akademisk hederlighet, referenser skall behandlas med respekt och återges utan förvanskning. Den andra aspekten, som inte berörs i denna uppsats, är att minimera integritetsintrång, vilket är särskilt viktigt vid behandling av information rörande extra skyddsvärda personer, exempelvis minderåriga.

Forskning och forskare existerar inte i ett vakuum, den omgivande kulturen formar tillsammans med personliga upplevelser föreställningar om den värld vi lever i. En forskare måste därför vara medveten om denna bias för att inte resultatet främst färgas utifrån egna åsikter, att dessa också klart redovisas, detta gäller speciellt inom humaniora då den i hög grad grundas på värderingar (Molander 2003:166). En aspekt av detta är urvalet av material, uppsatsens fokus hamnar i ett västerländskt fokus utifrån valet att granska europeiska och nordamerikanska organisationer. Samma studie i exempelvis Australien eller Afrika hade troligen vinklats annorlunda, vilket gör att avgränsningar och urval bör motiveras för att verka för eftertanke hos både forskaren och läsaren. Då forskningen ofta bedrivs utifrån ett personligt intresse finns ofta också en förförståelse vilken styr riktningen på arbetet, att därför tidvis stanna upp och reflektera över de egna resultaten bör vara en integrerad del av arbetsprocessen (Pripp 2011:70).

2 - Algoritmer och artificiell intelligens

Ur ett rent tekniskt perspektiv kan algoritmer ses som en uppsättning instruktioner, sortering och filtrering är exempel på abstrakt beskrivningar av funktioner för behandling av data. De kan även ses som en del av en digital vardag, den infrastruktur vilken styr, definierar och möjliggör mycket av dagens mänskliga aktivitet. De existerade också som idé redan innan datorernas intåg, en idé om tänkande maskiner vilka vi känner igen från litteraturen och filmens värld. Artificiell intelligens är det begrepp som framför allt behandlas i denna uppsats då policydokumenten använder detta, diskussionen kring de etiska förhållningssätten bör dock inbegripa även annan algoritmisk användning.

2.1 Artificiell (Generell) Intelligens

Artificiell intelligens, ett suggestivt och svårfångat koncept med en mängd konnotationer. Ett begrepp tidigare associerats till science fiction, robotar och framtida dystopier gestaltade i filmer och litteratur, vilket idag i allt högre grad används för att beskriva produkter och fenomen i samtiden. Ett närbesläktat begrepp vilket sällan förekommer i populärkulturen men är vanligare hos de som utvecklar programvaror för banksystem, medicinsk analys och industrier är *maskininlärning*, denna form av upplärda expertsystem blandas ofta ihop med just artificiell intelligens.

I diskussionen om algoritmer finns det i grunden två begrepp att hålla reda på, artificiell intelligens (AI) och artificiell generell intelligens (AGI), dessa har båg i varierande grad beröringspunkter med det denna uppsats avhandlar. Artificiell intelligens diskuteras ofta i media tillsammans med robotar, både i form av hot mot anställningstrygghet och något som skall förenklar allas vår vardag. Detta kan tyckas vara en paradox, vilken troligen beror på att artificiell intelligens, liksom så många andra verktyg, kan användas för bra såväl som mindre bra ändamål (Intel 2017:1-2).

När artificiell intelligens förekommer som en del av film och science-fictionlitteratur är oftast den korrekta benämningen artificiell generell intelligens. Detta är maskiner, som även när de inte försöker ta över världen, oftast har möjlighet att självständigt skapa förståelse för och anpassa sig till den. De maskiner som förväntas ta över människornas arbetsuppgifter drivs oftast inte av artificiell intelligens utan av maskininlärning. De är specialiserade på samma sätt som en brödrost endast rostar bröd, ett system för analys av röntgenbilder är idag överlägsna läkarna på att upptäcka cancerceller, men saknar kunskap om och förståelse för vad sjukdomen är och hur den kan botas (ibid).

2.2 Historisk tillbakablick

Det moderna konceptet för artificiell intelligens kan tillskrivas Alan Turing, matematiker och filosof vilken var verksam under första halvan av 1900-talet. Han har bland annat gett sitt namn åt begreppen Turingmaskin och Turingtest. De datorer som fanns under hans levnadstid var specialiserade på enstaka

uppgifter, ofta militära appliceringar som att räkna ut projektilbanor eller dekryptera koder. Turingmaskinen var en teoretisk modell för hur en dator programmässigt kan anpassas för generella uppgifter, med andra ord så som de fungerar idag. Genom att förse en dator med instruktioner tillsammans med den data som skall behandlas skapas en maskin med möjlighet att förändra den egna programkoden. Under en föreläsning 1947 menade Turing därför att en dator med ett tillräckligt stort minne därför borde kunna modifiera hur den löser sin uppgift, och genom detta bli effektivare än då den konstruerades, uppvisar den då inte intelligens? (Michie 2008:61-65).

Turingtest är ett grundläggande begrepp inom AI-forskningen, något som även speglar frågan om vad mänsklig intelligens är. Enkelt uttryckt hävdar Turing att när en människa inte längre kan avgöra om det är en annan människa eller en maskin den har en dialog med, är maskinen att anse som intelligent (Goertzel, Pennachin 2007:8). För att utföra detta test menade Turing att kommunikationen endast behöver vara i textform, exempelvis skulle då skeppsdatoren HAL i filmen *2001: Space Odyssey* passera som intelligent. Ett utökat test där utfrågaren även har tillgång till visuell information om hur varelsen finner sig tillrätta i världen kallas *total Turing test*, Roboten som skickas tillbaka från framtiden i *Terminator* får representera detta, en maskin vilken skulle vara svårt att skilja från en tänkande människa (Nepolitan, Jiang 2018:2).

Under de följande 60 åren efter Turing presenterat testet, har varken forskare eller ekonomiska medel saknats för att konstruera en maskin som uppfyller Turings krav. Det har dock visat sig vara komplicerat att skapa en modell av den mänskliga hjärnan, vare sig simulering av dess kognitiva processer eller råstyrka med större och snabbare datorer har visat sig vara framgångsrik. Det har inte heller saknats kritiska röster mot testets premisser, även om en maskin skulle upplevas som intelligent ses den av många filosofer fortfarande som en simulering och inte en intelligent tänkande varelse med medvetande (Nepolitan, Jiang 2018:4-7).

2.3 Litteraturens bidrag

Redan Mary Shelleys *Frankenstein; or the modern Prometheus* från 1818 kan ses som ett tidigt försök att litterärt utforska vad det är att vara människa genom att skapa artificiell intelligens. Ett verk som senare återberättats med olika grader av teknologiska framsteg, från genetisk manipulering till autonoma robotar. Inte sällan har detta dystopiska verk kombinerats med en determinism, hur den oundvikliga utvecklingen förslavad mänsklighet under dess egen skapelse (Lovén 1999:22).

När algoritmer presenteras i litterära sammanhang är det oftast i form av artificiell intelligens, allt från den deprimerade roboten Marvin i *Lifiarens guide till galaxen* skriven av Douglas Adams till Philip K. Dicks replikanter i *Do androids dream of electric sheep?* vilken för de flesta nog är mer känd genom Ridley Scotts filmatisering *Bladerunner*. Utopier och dystopier som dessa berör frågan om vad det är att vara människa utifrån hur vi behandlar *den andre*, trots uppvisandet av känslor i bägge dessa fall

behandlas dessa som maskiner. Det framstår även som om dagens normer vad gäller artificiell intelligens till viss del är baserad på de tidiga litterära verken, en rädsla att förlora kontrollen i kombination med förväntningar på evolutionära framsteg.

2.4 Artificiell intelligens - hot eller möjlighet?

Som inledningen av detta kapitel antyder är den artificiell intelligens som används idag inte densamma som *artificiell generell intelligens*, utan något betydligt enklare och mer specialiserat på enstaka uppgifter. Det är inte heller det som benämns *artificiell super intelligens* (ASI), något forskare som Nick Bostrom och Max Tegmark varnar för. AI-tekniken i sig beskrivs av Tegmark som neutral, hur vi väljer att använda den bestämmer om den blir ond eller god. De ser dock bägge en fara i hur dessa kommande superdatorer i en inte allt för avlägsen framtid kan komma att uppnå medvetande och på sikt helt konkurrerar ut mänskligheten (Lenas 2017). Teknikens neutralitet kan dock ifrågasättas, historieforskaren Melvin Kranzberg publicerade 1986 sina sex lagar om teknologi, varav den första lyder:

Technology is neither good nor bad; nor is it neutral.

Teknik sedd ur ett historiskt perspektiv skiljer sig från Tegmarks mer pragmatiska syn på teknikens användning, Kranzberg menar att teknik alltid färgas av den diskurs vilken den befinner sig i, vilket gör att synen på den också kommer att förändras med tiden (Kranzberg 1986:545-548). Inte bara den historiska diskursen kan färga hur vi bedömer nya tekniska tillämpningar, dess återverkningar kan även skilja utifrån geografiska och kulturella skillnader. Kranzberg påminner oss slutligen i sin sjätte lag att tekniken är en väsentlig del av mänsklighetens historia och att teknik helt saknar värde utan mänsklig interaktion:

Technology is a very human activity-and so is the history of technology.

Maskiner i sig är endast livlösa ting, Kranzberg menar att det ändå finns en uppfattning att dessa ting styr våra liv, klockan–inte tiden, styr vår vardag. Bilen och andra transportmedel kan begränsa vår rörlighet–inte bara socioekonomiska möjligheter som ekonomi och medborgarskap. Trots tekniska landvinningar finns det i grunden andra faktorer som styr vår vardag, det finns aktiviteter vi genomför på ineffektiva vis för att det är mänskligt. Videokonferenser har länge varit ett alternativ till fysiska möten, ändå väljer många att resa världen runt för att träffas personligen. Mötets värde ligger många gånger i den sociala kontexten och inte i dess dagordningen, inlägg och beslut (Kranzberg 1986:557-560).

Artificiell intelligens, hot eller möjlighet? Svaret blir, ur Kranzbergs synvinkel, att det beror på. Han menar att ur en historisk synvinkel har den tekniska utvecklingen varit gynnsam för mänskligheten i

stort. Med det sagt, det går också att se hur påverkan från en och samma teknik kan variera stort hos olika grupper. Som exempel nämner han användningen av DDT vilket förbjöds i stora delar av världen för sin negativa inverkan på ekosystemet, men den fortsatta användningen i Indien sparade många liv då den starkt minskade förekomsten av malariamyggor (ibid 546). På liknande sätt kommer troligen vissa att drabbas och andra att gynnas av AI-teknologin, ingen kommer dock lämnas opåverkad.

2.5 Algoritmer skapar förutsättningar för en ny ekonomi

Ända sedan persondatorer började sitt intåg, först på kontoren och sedan i hemmen har användningen styrts av den tillgängliga programvaran. Mängder av mjukvaruföretag utvecklade mängder med olika tillämpningar för att möta marknadens behov, några lyckades och många misslyckades, men så länge programvaran var välskriven fungerade den på majoriteten av de tillgängliga datorerna. Innan Internets framväxt kunde det dock vara svårt för mindre utvecklare att nå kunderna, till skillnad från mjukvaruföretag som Microsoft, Oracle och Electronic Arts med lokalkontor i flera länder var de beroende av externa distributionsföretag för sin försäljning. Internets intåg förändrade i ett slag detta då den globala marknaden kunde nås digitalt, kunder behövde inte längre köpa en kartong med ett fysiskt medium. I praktiken har dock de tidigare distributionsföretagen idag ersatts av distributionsplattformar, plattformar vilka är en integrerad del av de digitala enheter de flesta använder i sin vardag. Genom dessa plattformars inbyggda funktionalitet kan inköp av både programvaror, underhållning och fysiska produkter göras. Detta passar bra in i den beskrivning forskarna Jonas Andersson Schwarz och Stefan Larson benämner *digitala plattformar*, vilka de ser som ”en mjukvarubaserad, centralstyrd yta där olika tredjepartsaktörer kan mötas och finna varandra för olika typer av marknadsutbyten” (2019:5).

I västvärlden kontrolleras idag majoriteten av vår dagliga digitala användning av fem företag, Apple, Google, Facebook, Amazon och Microsoft, dessa benämns ofta *Big five* och verksamheten på dessa globala plattformar styrs från företagens huvudkontor i USA. Kina står i detta sammanhang ut som ett specialfall, där dominerar fyra företag, Alibaba, Baidu, Jingdong Mall och WeChat, de har verksamhet vilken i stort sett är begränsad till Kina och kontrolleras ytterst av den Kinesiska staten (Gelin. 2018:8).

Gemensamt för alla dessa plattformar är deras beroende av algoritmer och den enorma mängd data de bearbetar för att skapa sig en bild av sina användare. Datan är dels insamlad användardata vilken automatiskt genereras av applikationerna i form av till exempel GPS-data, men även den vi användare frivilligt lämnar ifrån oss när vi till exempel söker efter ett matställe i TripAdvisor. Hur all denna data sedan används bestäms utifrån plattformens affärsmodell och begränsas av användaravtalets samtycke (Van Dijck, Poell, De Waal 2018:9).

Artificiell intelligens är idag något som används för att sälja såväl kylskåp¹ som banksystem², och även om kylskåpet varken förstår innebörden av vad du säger eller ens gör vardagen lättare, blir ändå teknik som röststyrning ett allt vanligare försäljningsargument. När det blir möjligt att kommunicera med fler av våra maskiner på detta vis kan det givetvis bli svårt att hålla reda på vilken maskin som avses. ”Hej Siri” och ”Hej Google” kan då vara lösningen, en plattform som styr de andra maskinerna. En produkt som blir både personlig betjänt och övervakare från ett och samma företag.

¹ https://www.hoover.co.uk/en_GB/axi-refrigerator

² <https://www.sparkbit.pl/artificial-intelligence-banking-part-2/>

3 - Sammanställning av policydokumenten

Analysen av policydokumenten kommer, vilket nämnts i metodavsnittet, att genomföras med metoden *what's the problem represented to be* (WPR). Först presenteras här en sammanfattning av de fem policydokumenten, framför allt områden där det kan förekomma etiska implikationer.

3.1 Offentliga aktörer

De tre offentliga aktörerna representerar EU, Storbritannien och USA och spänner över åren 2016 till 2018. Dokumenten från Storbritannien och USA är skrivna av rådgivare och det tredje är framtaget av EU-kommissionen för att förmedla sin syn på artificiell intelligens till övriga organ inom EU.

3.1.1 Europeiska kommissionen

I sin kommunikation med övriga organ inom det Europeiska samarbetet meddelade kommissionen 2018 sin syn på algoritmers betydelse för framtidens union i dokumentet *Artificial Intelligence for Europe*. Där slår de redan i inledningen fast att artificiell intelligens har lämnat fiktionens värld och i högsta grad redan är en del av vardagen. I inledningen anas även ett i det närmast förbehållslöst omfamnande av den nya tekniken, som en kraft för ökat välstånd och välbefinnande.

En genomgående trend i dokumentet är diskussionen om EU som relevant deltagare i kampen om global dominans, investeringar för att öka kunskap och kapacitet anges som en måttstock på detta. Konkret anges även behovet av moderniseringen av utbildning för att möta en ny arbetsmarknad, underförstått tycks modernisering vara detsamma som ökad fokus på digital teknik. Slutligen ser kommissionen ett behov av etiska riktlinjer, utarbetade i ett samarbete mellan samtliga intressenter och baserat på EUs grundläggande rättigheter (European Commission 2018).

3.1.2 Executive Office of the President (of the United States)

Denna rapport, författad under Barack Obamas presidentskap, är skrivet av rådgivare från flera delar av administrationen. Den bygger framför allt på ekonomisk såväl som teknisk kunskap men står även med ett tredje ben stadigt i inrikespolitiken. I likhet med sina europeiska kollegor inleder de med ekonomiska aspekter på teknologin, hur den idag är den viktigaste kuggen i motorn vilken driver den ekonomisk utveckling framåt, viktigare än tillverkning såväl som kapital (s. 1). Samtidigt som diskussionen om denna fjärde industriella revolution vilken de benämner som den mest omvälvande i mänsklig historia, uppmärksammas även en inbromsning i en ökad global produktion (s. 7). De höjer därför ett varningens finger för att en ökad automation inte skapar ökat välstånd för alla, de nya jobben som skapas förväntas inte kunna ersätta alla de som förloras.

Rapporten presenterar förslag till policy genom tre konkreta strategier, även här anas syftet att behålla hemmamarknaden USA som en relevant medtävlare på den globala arenan.

1. Invest In and Develop AI for its Many Benefits

Ökade investeringar, i såväl teknologi som forskning, ses som en garant för USA som ledare inom artificiell intelligens, där ett strategiskt övertag ses som ett mål i sig.

2. Educate and Train Americans for Jobs of the Future

Liksom Europeiska Kommissionen tar rapporten fasta på behovet av utbildning inom digital teknik, naturämnen och matematik, här räknar de även upp humanioras specialområde, kritiskt tänkande, som betydelsefullt för en framtida arbetsmarknad.

3. Aid Workers in the Transition and Empower Workers to Ensure Broadly Shared Growth

Denna rapport är framtagen under Obamas presidentskap, vilket kan förklara förekomsten av förslag på stärkta sociala skyddsnät under den övergångsperiod då många från socioekonomiskt utsatta grupper förväntas bli friställda. Dessa är åtgärder för sjukvård, utbildning och ersättningar till arbetslösa som närmast för tankarna till europeiska välfärdssystem.

3.1.3 Government office for science (Storbritannien)

Denna brittiska rapport från 2016, strax innan Brexit var ett faktum, är författad av den statlige vetenskapsrådgivaren Mark Walport. Den fokuserar i huvudsak på den statliga användningen av artificiell intelligens och dess effekt på den brittiska befolkningen i allmänhet. Liksom de tidigare rapporterna förutser Walport stora omställningar på arbetsmarknaden men presenterar samtidigt en oenig samling experter där hälften av de tillfrågade anser att satsning på automation och artificiell intelligens kommer att öka antalet arbetstillfällen, den andra hälften att de kommer att minska (s. 12). Detta om något illustrerar svårigheten med att förutsäga vilket genomslag den nya tekniken förväntas få, denna oenighet om dess effekter visar även på behovet av etiska riktlinjer för att om möjligt förekomma de problem som kan uppstå, vilket också framkommer i rapporten (s. 14).

De teman som framträder i rapporten kan sammanfattas till ett behov av ökad användning av artificiell intelligens i myndighetsutövning, dess återverkning på arbetsmarknaden samt hur allmänhetens acceptans och förtroende för den nya tekniken kan uppnås genom dialog och transparens. Walport ser stora vinster med ökad automatisering av beslutsstöd med hjälp av artificiell intelligens och hur vården kan effektiviseras och samtidigt erbjuda kvalitativt bättre vård med hjälp av nya analysverktyg. De arbetsuppgifter som förutspås förändras eller försvinna är framför allt rutinjobb, detta gäller både arbetare och tjänstemän. Däremot kommer, enligt rapporten, efterfrågan på högutbildad personal att vara större än tillgången, speciellt i yrken som är svåra att automatisera där kreativitet och social kompetens är viktigt. Generella kompetenser kommer, enligt Walport, därför att bli allt viktigare för anställningsbarhet (s. 13).

3.2 Privata aktörer

De två privata företagen är bägge från USA men är utvalda främst för sin globala och i flera avseenden dominerande ställning. IBM med sina dryga 100 år i branschen och sin dator *Watson* får tillsammans med Intel, multinationell processortillverkare med en tidigare närmast monopolställning, får representera industrins inställning till utvecklingen av artificiell intelligens. Bägge dessa dokument är publicerade 2017.

3.2.1 IBM Cognitive Principles

Som en av de ursprungliga giganterna i datorindustrin har de onekligen inflytande på utvecklingen av artificiell intelligens. Redan i inledningen av detta dokument beskrivs hur de ser artificiell intelligens som en transformativ kraft på hela samhället, och hur de med stöd av mer än ett sekels erfarenhet av ny teknik, ser en nödvändighet av att tidigt i utvecklingen fastställa riktlinjer för vilka tillämpningar som skall utvecklas och hur de skall införas. I inledningen utgår de även från premissen att artificiell intelligens radikalt förbättra både yrkes- och privatlivet utifrån den nya teknikens förmåga att bearbeta allt större mängder av data i allt högre hastighet. Som riktlinjer för sin framtida verksamhet slår de fast tre principer för vad de benämner ”den kognitiva eran”.

1. Purpose

IBM anser inte att artificiell intelligens har något egenvärde då de anser att ett sånt system inte kan uppnå det vi benämner medvetande. Istället är teknikens syfte att vara en förstärkning av- och utöka människors förmågor, hela tiden under mänsklig kontroll.

2. Transparency

IBM avser att klargöra sin användning av artificiell intelligens på tre sätt, detta för att öka tilltron till det slutresultat processen syftar till.

- När och till vilket syfte AI-teknik används.
- Tydlighet i hur algoritmerna är framtagna, dess bakgrundsdata och kompetenser.
- IBM erkänner sina kunders äganderätt till sin data och sina affärsmodeller, de erbjuder artificiell intelligens som ett sätt att öka sina kunders potential på marknaden.

3. Skills

Artificiell intelligens har, enligt IBM, ingen plats i samhället om inte hänsyn tas till på vilket sätt maskinerna skall samarbeta med människor, hur de på bästa sätt hjälper mänskligheten att uppnå sin fulla potential i den kognitiva ekonomi som växer fram.

3.2.2 Artificial Intelligence; The Public Policy Opportunity (Intel)

Detta dokument från processortillverkaren Intel sticker ut från mängden då det är ett globalt företags rekommendationer för hur offentlig sektorn bör utveckla sin policy angående artificiell intelligens. De inleder med att slå fast att medborgare knappast litar på artificiell intelligens om utvecklingen styrs endast utifrån kommersiella intressen, därför ser de ett behov av statliga kontrollorganisationer. De är även fast övertygade om att användningen av artificiell intelligens har enorma möjligheter att förbättra livet för mänskligheten och att syftet med policyns utformningen är att underlätta utvecklingsarbetet. Deras rekommendationer presenteras i fem punkter.

1. Foster Innovation and Open Development

Att främja forskning och utveckling inom artificiell intelligens innebär enligt Intel att främja forskning som leder till en ökad användning av och nya områden för AI-tekniken. De förespråkar samarbete mellan offentliga organ, skolor och det privata näringslivet för att på alla nivåer föra in medvetenhet om AI-teknikens möjliga användningsområden. De ser gärna att ungdomar uppmanas att söka sig till utbildningar vilka leder till kompetens inom artificiell intelligens och att lärosäten i högre grad prioriterar utbildningar som leder till utveckling av morgondagens artificiella intelligens.

2. Create New Employment Opportunities and Protect People's Welfare

En högre grad av automatisering med hjälp av artificiell intelligens kommer att påverka arbetsmarknaden, framför allt positivt genom effektivisering men även negativt då hela yrkeskategorier kan komma att försvinna till följd av denna automatisering. Det som då behövs, enligt Intel, är att utarbeta hur samma teknologi - artificiell intelligens, kan användas för att hjälpa de drabbade till nya anställningar.

3. Liberate Data Responsibly

Inom offentlig sektor skapas idag enorma mängder med data om medborgare, databaser med medicinska, offentliga och personlig uppgifter vilka med hjälp av AI-teknik, förutom nya botemedel även kan leda till lösningar på andra samhällsproblem. Fri tillgång till dessa data beskrivs som oproblematiske så länge som den hanteras säkert i förhållande till den personliga integriteten. Inget nämns i sammanhanget om vem som äger resultatet utifrån behandling av denna fria data.

4. Rethink Privacy

Rekommendationen för att öka säkerheten runt privat data är uppdelad i två delar, den första lösningen bygger på samarbete mellan säkerhetsföretagen och systemägarna. Använd artificiell intelligens för att förutsäga var cyberattacker sker, med nog stort dataunderlag bör en artificiell intelligens kunna tränas i att upptäcka intrång innan det hinner fullbordas. Intel förespråkare även kraftfulla regleringar av säkerheten runt AI-tillämpningar såväl som andra system vilka hanterar identifierbar personlig information. För att uppnå detta förespråkar de vad som benämns *Privacy by*

design, vars innebörd är att teknikutvecklare redan i planeringsstadiet specificerar hur personlig information skall hanteras och säkerställas.

5. Require Accountability for Ethical Design and Implementation

Det är viktigt att de företag och organisationer som nyttjar artificiell intelligens också kan uppvisa att detta sker ansvarsfullt. Med detta menar Intel att alla nivåer av företaget/organisationen konsekvent följer de etiska riktlinjer som gäller internt och externt, viktigt är också att de förmår kommunicera detta till såväl kontrollinstanser som allmänheten. Transparens är ett område där framför allt privata företag kan få problem, dels för att företag traditionellt vill hålla strategisk information hemlig men även för att företagen själva inte alltid vet hur algoritmernas resultat producerats.

3.3 Sammanställning av dokumenten

Gemensamt för både de privata företagen och de offentliga organisationerna, är synen på behov av transparens för att erhålla förtroende från medborgarna. De ser detta som nödvändigt för en ökad spridning och användning av tekniken. Ett tema som står ut i de offentliga organisationerna är kampen om att leda utvecklingen, det framstår som om en ledarposition inom artificiell intelligens garanterar ett ekonomiskt försprång på en global arena. De är alla även överens om AI-tekniken som transformativ, mestadels ses även positivt på denna förändring. Det fortsatta analysarbetet utgår därför från denna sammanställning genom tre teman:

- Regler och transparens krävs för att artificiell intelligens skall nå sin fulla potential.
- Automatisering och artificiell intelligens är en kapprustning med allt högre insatser.
- Satsning på artificiell intelligens skapar nya kreativa yrkesroller istället för de som förloras.

4 - Analys

Innan analysen följer en förklaring av hur de sex stegen enligt Bacchis metod används i denna uppsats, analysen är sedan uppdelade efter de tre teman som framkommit i policydokumenten enligt 3.3.

4.1 Vilka problem representeras?

Genom de två första stegen i WPR-metoden analyseras vilken problematisering som rekommendationen är tänkt att adressera, då avslöjas enligt Bacchi den bakomliggande uppfattningen som förmedlas i dokumentet³. Genom detta sätt att arbeta sig bakåt, från förslag till ursprungligt problem, går det i senare steg att analysera om förslaget är det enda sättet att angripa problemet (Bacchi 2009:3). Uppfattningarna i analysen utgår från kulturella värderingar, och då samtliga dokument är tagna ur en västerländsk diskurs är denna på ett övergripande plan tämligen homogen. Ideologier på områden som privat ägande och personlig integritet skiljer säkerligen något, men markanta skillnader uttrycks inte i de analyserade dokumenten. En ökad förståelsen för hur uppfattningarna skapas kan hämtas genom en analys av diskursen, vilken i detta fall är de organisationer som tagit fram dokumenten. Genom att se hur dikotomier i texten värderas, hur centrala koncept definieras, och slutligen kategoriseringar används för att skapa mening, går det att få en uppfattning om vilka föreställningar som rekommendationerna bygger på (ibid 7-9).

Tredje steget i analysen är att undersöka dels vilka faktorer som är bidragande till uppkomsten av nämnda problemställningar, men om möjligt även att hitta ursprunget till problemet. Bacchi menar att upprinnelsen till att något ses som ett problem går att finna både historiskt–problemet kunde blivit annorlunda, och utifrån specifika beslutsprocesser (Bacchi 2009:10).

Genom att i det fjärde steget undersöka vilka aspekter av algoritmer som inte omnämns går det ofta, enligt Bacchi, att granska problemställningar ur andra perspektiv, vilket också ifrågasätter den ursprungliga problematikens underliggande premisser (Bacchi 2009:12-13). Om fördelningspolitiken i världen representeras som ett problem utifrån extrem rikedom istället för extrem fattigdom förändras också hur vi ser på de bakomliggande orsakerna liksom lösningarna.

Femte steget undersöker vilka effekter problemrepresentationen kan få, inte bara på själva diskursen utan även på de personer och grupper som är direkt berörda. Det går även att fundera på vilka förändringar som kan förväntas, vilka som är vinnare och förlorare och om det finns någon trolig syndabock (Bacchi 2009:15-18).

Sista steget i analysen knyter tillbaka till det tredje, uppkomsten av problemrepresentationen. Vad eller vilka står bakom synen på algoritmer som denna naturlig del av hela den framtida mänskliga kulturen,

³ Bacchi understryker att det varken är den allmänna uppfattningen eller policy-författarnas inställning som analyseras i detta steg, utan vilka förutfattade mening som kan utläsas ur textens "problem" (Bacchi. 2009. s.5).

vilka grupper är det som har möjlighet att forma diskursen värden och vilka motsätter sig dessa? Detta steg analyseras gemensamt för alla tre teman i avsnitt 4.5.

4.2 Regler och transparens krävs för att artificiell intelligens skall nå sin fulla potential

Något de flesta verkar vara överens om är betydelsen av allmänhetens tilltro till tekniken, Mark Walport skriver under rubriken "Public trust" (Government office for science 2016:17):

Public trust is a vital condition for artificial intelligence to be used productively. Trust is underpinned by trustworthiness. But whilst this can be difficult to demonstrate in complex technical areas like artificial intelligence, it can be engendered from consistency of outcome, clarity of accountability, and straightforward routes of challenge.

Denna lydelse är representativ för hur dokumenten i allmänhet ser på betydelsen av att allmänheten tar till sig och använder tekniken. Fokus ligger på två koncept, tilltro till tekniken behövs för att artificiell intelligens skall kunna användas produktivt, vilket bygger på att processen ska kunna följas upp, någon skall kunna ställas till svars.

För den potential som förväntas uppnås förekommer två kategorier, dels teknikens möjlighet att skapa en behagligare vardag och främja en bättre hälsa hos mänskligheten. Det finns även dess ekonomiska potential, hur införandet av automatisering genom artificiell intelligens effektiviserar tillverkning och skapar helt nya marknader. Underförstått är verksamhet understödd av artificiell intelligens att föredra, så länge som hänsyn tas till det Walport nämner sist i citatet ovan. Artificiell intelligens får inte bli en svart låda vilken levererar godtyckliga svar, även om systemens komplexitet gör processen närmast omöjliga att överskåda måste resultaten fortfarande gå att ifrågasätta. Detta är ett problemområde Intel anser kan komma att lösas med hjälp av AI-tekniken själv, algoritmer som kan förklara sitt resonemang är under utveckling, likaså forskning på artificiell intelligens som själv kan visa vilka regler som använts i processen (Intel 2017:9). Transparens tas upp i flertalet av dokumenten som en nödvändig del av relationen mellan organisationer och allmänheten, alla tar dock inte upp problematiken att företag sällan är villiga att avslöja sina processer, då de ger dem ett övertag på en konkurrensutsatt marknad. Transparens och regler går också hand i hand, utan regler uteblir bedömningen av det som ses, och utan transparens blir uppföljningen svår att genomföra.

Att AI-tekniken skall uppnå sin fulla potential kan komma i konflikt med föresatsen att bekämpa bias hos algoritmerna, det kan till exempel mycket väl vara så att det mest effektiva sättet att bekämpa brottslighet är att använda data om nuvarande dömda, trots risken att denna data kan bygga på både rasism och fördomar. Trots detta finns en uttalad förhoppning i flera av dokumenten att tillgång till större mängder data kommer att göra det möjligt att på ett bättre sätt träna upp algoritmerna.

Det underliggande problemet kan tydas som att medborgarna inte litar på resultaten producerade av artificiell intelligens och att det saknas regler för hur artificiell intelligens får användas i tillämpningar.

Idé- och lärdomshistorikern Sven-Eric Liedman använder sig av begreppet ”modernitetens förtrollning” för att beskriva samtidens acceptans av en allt snabbare digitalisering av samhället (1997:454). Att det är en förtrollning baserar han på den föreställning som finns om framtiden, hur den främst är baserad på förhoppningar snarare än rationellt tänkande (ibid 461). En titt i backspegeln visar också på hur tekniken gång på gång förändrat samhällsutvecklingen. Ända sedan 1700-talet och upplysningstiden har vetenskapen erbjudit nya sätt att förstå vår värld, ny kunskap skapat förutsättningar för avancerade maskiner, först som ersättare våra muskler och senare vårt tänkande. Varje ny epok möttes troligen av både fascination och rädsla, ångmaskinen, glödlampan och televisionen, alla motsvarar de magi och teknik i ett och samma objekt. Dessa ofta samhällsomvälvande tekniker har historiskt både reglerats och finansierats av staten, el- telefon- och transportnät, infrastruktur vilken är allmännyttiga och därför bör skyddas mot möjligheten till missbruk genom monopolställning. Även om det finns undantag så är fortfarande Internets underliggande struktur till stor del kontrollerad av den icke vinstdrivna organisationen ICANN⁴, att rent praktiskt söka information, nå ut med ett budskap eller kommunicera med andra är dock mestadels kontrollerat av ett fåtal globala företag och deras algoritmer (Dijck, Poell, Waal 2018:12ff).

För att studera problemet med tillit går det att dra paralleller med till exempel säkerhetspolisens arbete, även här är det få personer som har insyn. Även utan statistik går det att ana att befolkningen, åtminstone i Sverige, ändå i hög grad har förtroende för deras arbete. Så om inte transparens ensamt är problemet, kan det handla om verksamhetens grundläggande motiv? Säkerhetspolisens uppgift kan ses som att avstyra hot mot landets säkerhet, dess demokratiska principer, Googles och Facebook liksom de flesta företag bör istället ha som mål att öka sina aktieägares förmögenhet. Ekonomiska motiv är dock inte det enda sättet att nå framgång, systemen behöver inte vara slutna. På Internet finns ett fenomen som tydlig avviker från de övriga plattformarna, Wikipedia. Som plattform kan den ses som ett försök att infria det tidiga internets löfte, en mötesplats där alla kan göra sin röst hörd och ta del av andras kunskap. Skaparen själv, Jimmy Wales, anser att Wikipedias uppdrag är att ge världens befolkning tillgång till all mänsklig kunskap, och alla som har möjlighet kan dela med sig (Isaacson 2014:497). Här finns inga aktieägare, inga monetära vinstintressen och inga annonser, organisationen finansieras helt av donationer. Istället för att övertyga befolkningen om att lämna över sina personliga uppgifter till kommersiella aktörer som Google och Bing i utbyte mot tillgång till en idag så grundläggande tjänst som sökning, kanske världssamfundet bör samarbeta om en gemensam öppen plattform för detta. Det kommer fortfarande inte gå att arbeta med full transparens, då ett öppet system lättare kan missbrukas.

⁴ Internet Corporation for Assigned Names and Numbers är den organisation som sköter driften av de centrala servrar som håller reda på internets kataloger över domännamn.

Men utan ett ekonomiskt vinstintresse bör det rimligen finnas större förutsättningar att motverka bias, något som kan öka förtroendet för tekniken (Nobel 2018:152).

Frågan är dock om det går att lösa frågan om bias och etiska frågeställningar trots transparens och publika plattformar så länge som majoriteten av de verksamma inom fältet består av vita män. En artikel i *The Guardian* redovisar hur snedfördelad personalsammansättningen är på arbetsplatser som Facebook, Microsoft och Google, när endast några enstaka procent är färgade och upp till 80 procent är män är risken för oavsiktlig bias i system stor. För att åstadkomma en förändring av denna obalans måste företagen och samhället aktivt arbeta för detta. En svårighet med personalsammansättning som artikeln pekar på är tillgängligheten till utbildad arbetskraft, att i USA är det mer än dubbelt så många män i förhållande till de kvinnor som söker arbeten relaterade till artificiell intelligens (Paul 2019).

Möjligen bör vår framtida tillit till artificiell intelligens jämföras med en annan teknisk landvinning vilken vi sällan reflekterar över idag, elektriciteten. Vi använder den dagligen, förlitar oss till att den ström som finns i uttagen är av rätt typ, gärna miljömärkt, men ändå vet vi att fel använd är den livsfarlig. Vi kan absolut använda artificiell intelligens för mänsklighetens bästa på en mängd områden som hälsa och farliga arbetsmoment, men samtidigt vara medveten om riskerna och inte plugga in den där den inte hör hemma.

4.3 Automatisering och artificiell intelligens är en kapprustning med allt högre insatser

Policydokumentet från EU-kommisionen uttrycker i klara ordalag att utvecklingen och användningen av artificiell intelligens är en kapprustning, där andra delar av världen ökar sina investeringar, och Europa riskerar att ”missa tåget” (European Commission 2018: Kapitel 2). Behovet av ökade investeringar, privata såväl som offentliga, rekommenderas av samma anledning även av de amerikanska och brittiska rådgivarna. Att det är en tävling inbegriper såväl vinnare som förlorare, där finns också underförstått en vinst värd att kämpa för. Vad denna vinst består i finns inte direkt uttalad, mycket av argumentationen rör dock effektivisering i produktion av varor och tjänster, vilket bör leda till ökade vinster för de inhemska företagen och en starkare ekonomi för den berörda regionen.

Stöd för ökad användning av artificiell intelligens förväntas också att resultera i nytta, i den rapport som togs fram under president Obamas administration anges som första rubrik ”Invest In and Develop AI for its Many Benefits”. Stycket som följer blir ett cirkelresonemang som avslutas med att den viktigaste nyttan (benefit) med ökade investeringar är att bibehålla det övertag de då redan ansåg sig ha. Premissen som kan utläsas är att investeringar och utveckling av AI-tekniken per automatik kommer att medföra fördelar, större investeringar ger större vinster och den som investerar mest står som segrare (Executive Office of the President 2016:27).

Även Europa anses, enligt EU-kommisionens dokument, vara världsledande inom AI-teknologi. Åter igen används begreppet nytta (benefit) "[T]he benefits of adopting AI are widely recognised", vilket sedan understöds med *ett* exempel genom att referera till en undersökning som kommissionen själv tagit fram. Detta återfinns i 2018 *Digital Transformation Scoreboard*⁵ och visar på att företag inom livsmedels- och byggindustrin som tagit till sig AI-teknologi också uppvisar en positiv utveckling.

Intressant nog återfinns inte tävlingsinslaget när Intel presenterar behovet av ökad forskning och ökade investeringar, de räknar istället upp vinster för samhället. Det är inte läkemedelsindustrins utveckling som betonas utan läkarnas möjlighet att rädda liv genom bättre och snabbare diagnoser. Inte bilindustrins möjlighet till automatisering och ökad försäljning i och med självkörande bilar, det utelämnas till förmån för ökad trafiksäkerhet och minskad klimatpåverkan.

Den brittiska stadsapparaten ser dock hur införandet kan ge stora vinster hos den offentliga sektorn såväl som privata företag, de största vinsterna förutspås effektiviseringar stå för. De ser hur användning av big data försvårar fusk och därför ökar skatteintäkterna och hur företag som bättre använder kund- och konsumentdata också ökar sin produktivitet (Government office for science 2016:9). Denna rapport, skriven av den brittiska vetenskapsrådgivaren, ger uttryck för att Storbritannien är världsledande på AI-forskning, och menar att automatisering kommer att bli till stor nytta för hela landet.

Nytta (benefit) är ett begrepp som ofta återkommer i dessa resonemang, det nämns i svepande ordalag tillsammans med vård, hälsa och klimat. Detta förklarar dock inte behovet av att leda utvecklingen då speciellt klimatet inte erkänner några territoriella gränser, det gör däremot ekonomisk tillväxt och konkurrenskraft på en globala marknaden. Mycket talar för att kapprustningen till stor del handlar om att dominera den framväxande AI-baserade ekonomin, även då osäkerheten är stor om exakt vad detta i slutänden leder till.

Det finns en historia av det som Joseph Schumpeter benämner skapande förstörelse (creative destruction), när ekonomisk tillväxt sker genom att ny teknik tar över den föregående. Då denna nya form av ekonomi växer fram som resultat, tar den även över som drivande motor i tillväxten.

Industrialismens nationsbaserade serieproduktionen ersattes av ett globalt nätverk av tillverkare när Internet möjliggjorde detta, med den en viktlos ekonomi till stor del baserad på information (Miller 2011:47-48,64-71). Ett exempel på detta är musikindustrins försöka att anpassa sig till det digitala samhället genom CD-skivor, samma affärsmodell men med ny teknik. Detta omkullkastas i och med de digitala plattformarnas intåg, med streamingtjänster som Spotify faller den tidigare formen av distribution av fysisk media och samtidigt den tidigare ekonomiska modellen.

⁵ https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/Digital%20Transformation%20Scoreboard%202018_0.pdf

Artificiell intelligens kan antingen ses som en fortsättning på utvecklingen av denna informationsekonomi eller som en ny destruktiv kraft med möjlighet att stöpa om industrin ännu en gång. Klart är att nuvarande policy är baserad på en ekonomisk modell där effektivisering och öppnande av nya marknader ger ständigt ökad produktion och högre vinster. Detta i en tid där den då 15-åriga eleven Greta Thunberg⁶ strejkar från skolan, vilket blir startskottet för en global ungdomsrörelse som kräver en ny klimatpolitik. Underförstått innebär detta ett behov av lägre energiförbrukning och minskad konsumtion, detta är ett scenario som helt saknas i dessa policydokument.

Problemet kanske då inte är frågan vart industrin är på väg, utan istället en rädsla att bli kvar i ett system som inte längre kan hävda sig i den nya ekonomi som växer fram, vilken form den än tar. Kan detta vara anledningen till det som Andersson Schwarz och Larsson menar är EU-kommissionens verkliga mål med GDPR, att möjliggöra åtal mot plattformar som Google och Facebook? De skulle därigenom begränsa dessa plattformar, som uteslutande är amerikanska, för att istället gynna inhemska europeiska intressen, styra utvecklingen istället för att försöka följa den (Andersson Schwarz, Larsson 2019:373).

Var avregleringarna av kapitalmarknaderna och avskaffandet av mellanstatliga handelshinder att likna vid öppnandet av Pandoras ask? Istället för det dynamiska samspelet mellan företagen och staten under industrialismen försöker dagens globala företag som Apple istället undviker nationella lagar och regler genom avancerad planering. Länder som Irland lockar med fördelaktiga skattevillkor med följden att företag väljer att lägga sina Europa-kontor där, men produktionen ligger fortsatt i länder med lägre löner såväl som lösare regler på arbetsmiljöområdet. Detta är ett exempel på hur policy på olika områden krockar, där det inte går att tillgodose önskemålet om ökad produktion genom automatisering och effektivisering utan att göra avkall på andra principer. En sådan problematik är artikel 3 i EU-fördraget, i förbindelsen med omvärlden ska Unionen enligt femte stycket "[...]bidra till fred, säkerhet, hållbar utveckling av vår jord, solidaritet och ömsesidig respekt mellan folken, fri och rättvis handel, utrotning av fattigdomen och skydd för de mänskliga rättigheterna[...]". Att utveckla och införa ny teknik kostar, har små kongolesiska eller afganska jordbruk råd med mjölkrobotar, tillåter den afrikanska infrastrukturen självkörande bilar? Risker att en effektivisering av Europas jordbruk, produktion och transporter istället motverkar solidaritet, rättvis handel och minskad fattigdom, och därigenom ökar de socioekonomiska klyftorna i världen. För att nå de mål som nämns i policydokumenten, att forskning och utveckling behövs för mänsklighetens bästa, för att avvärja klimathotet, säkrare diagnoser i sjukvården och tryggare trafikmiljö, torde det vara lättare att nå dessa genom samarbete istället för protektionism.

Att AI-tekniken kommer på bred front och inom en snar framtid kommer att påverka många delar av vår vardag är dokumenten samstämmiga om, de är även överens om vikten av investering i forskning

⁶ <http://time.com/5463721/most-influential-teens-2018/>

och utveckling för att vara ledande på området. Till skillnad mot kalla krigets kapprustning där kontrollen låg hos de folkvalda, hamnar dock kontrollen av plattformarna och big data ofta i händerna på privata företag. Framför allt EU verkar vilja öppna upp den insamlade data som finns i offentliga register, artificiell intelligens behöver stora mängder av data och målet är att data skall flöda fritt inom Unionen liksom den idag fria rörligheten för varor, personer, tjänster och kapital. Att släppa denna data fritt tillgängligt skapar troligen stora möjligheter för innovativa entreprenörer att skapa nya tjänster. Detta även inom områden som hälsa och sjukvård då även denna data föreslås, i anonymiserad form, öppnas upp för användning. Balanserat mot detta finns förordningar som GDPR, den ger Unionens medborgare viss möjlighet att bli informerad om vilken information företag och organisationer använder och hur den bearbetas, samt krav på ett uttalat behov hos organisationen som vill använda personlig information. Intel betonar vikten av tillsyn från myndigheterna, en stor del av allmänhetens förtroende till artificiell intelligens anses bero på denna. Samtidigt ser de gärna fri tillgång till data för att maximera de positiva effekter AI-tekniken kan ge. Med synen på data som en råvara liknande skog, kol eller vatten går det att se behovet av regleringar av hur denna immateriella resurs får nyttjas, detta för att motverka risken för rovdrift (Van Dijck, Poell, De Waal 2018:141). Det är även viktigt att inte endast se större mängder data som ett mål i sig, kvaliteten på slutresultatet av algoritmernas bearbetning är även beroende av förståelsen av vad informationen representerar. Att utgå från att hälsodata i patientregister är representativt för en hel population ger ett snedvridet resultat, denna data speglar endast de som haft kontakt med sjukvården. Det är även viktigt att se till hur informationen samlats in, vilka uppgifter utelämnades, kvalitet är ofta viktigare än kvantitet (boyd, Crawford 2017:668-670).

Om målet är att framtidens digitala samhälle verkligen ska förbättra hela mänsklighetens situation behövs samarbete snarare än en kapprustning, det behövs gemensamma forskningsprojekt och multilaterala överenskommelser om etiska regler kring artificiell intelligens, allt detta för att i slutändan skapa goda relationer med morgondagens robotkollegor.

4.4 Satsning på AI skapar nya kreativa yrkesroller istället för de som förloras.

Det finns en konsensus i frågan om att satsning på artificiell intelligens leder till en omställning av hela arbetsmarknaden, även om den inte förväntas bli fullständigt genomgripande på alla områden (Executive Office of the President 2016:26). Inom såväl tillverkningsindustrin som tjänstemannasektorn förutspås repetitiva arbetsuppgifter försvinna, detta gäller såväl lågavlönade som mellanpositioner. De nya uppgifter som skapas kommer troligen att vara serviceyrken, ofta på samma arbetsplats men då med krav på högre digital kompetens (Government office for science 2016:12). EU-kommission har i sin tur stora förhoppningar på alla de nya arbetstillfällen som kommer skapas, de ser hur både äldre och funktionshindrade med stöd av ny teknik får möjlighet att vara en produktiv del av framtidens arbetsmarknaden (Kapitel 3.2).

AI-tekniken framställs från de offentliga beslutsfattarna som ett stöd för medborgare att bli effektivare och mer produktiva. Tekniken hjälper de anställda genom att underlätta tunga och farliga moment, de slipper monotona, repetitiva arbetsuppgifter. Detta gäller även tjänstemannasektorn där artificiell intelligens ses som framtida kollegor på kontoret, ett samarbete mellan personal och teknik där datorerna utför bearbetningar av stora datamängder och tar över rutinuppgifter. Att tekniken på detta sätt hjälper de anställda leder i viss mån till ett cirkelresonemang, anställda behöver stöd av artificiell intelligens för att vara attraktiv på en allt mer av artificiell intelligens automatiserad och effektiviserad arbetsmarknad.

Generellt sett ses en framtida arbetsmarknad som gynnsam, det är istället övergångsperioden som beskrivs som bekymmersam. Det finns ett fokus på att genom politisk styrning främja utbildningar vilka leder till ökad kompetens inom fälten som benämns STEM (Science, technology, engineering och mathematics). Detta ökade tillflöde av högre utbildade arbetare förväntas också lösa problem som bias hos artificiell intelligens då det antas öka diversifierad av anställda inom forskning och utveckling (Executive Office of the President 2016:29).

Framtidens arbetsmarknad framställs som om utmaningen först och främst handlar om digital kompetens, det finns en tilltro till att om bara de yngre väljer *rätt* utbildning kommer nya arbetstillfällen att skapas för dem. Premissen här framstår som att en automatisering genom artificiell intelligens förväntas effektivisera företagen globalt medans de nya kreativa, digitala arbetstillfällen kommer att skapas främst i Europa respektive USA.

Tidiga pionjärerna drivs ofta av nyfikenhet och idealism, det tidiga Internet var också en plats för många entusiaster som motiverades av att möjliggöra det omöjliga. Det var dock först när det kommersiella krafterna fick upp ögonen för den inneboende potentialen av ett världsomspännande informationsnätverk som också globaliseringen satte fart på allvar. Kapitalets behov av att hitta nya marknader fann i Internet det perfekta verktyget. Nationella restriktioner harmoniserades mellan stater för att inte utgöra handelshinder samtidigt som digital distribution minskade betydelsen av företagens geografiska placeringen (Miller 2011:93). Det går inte att förneka att den digitala revolutionen i grunden har förändrat människors livsvillkor, många gånger till det bättre. Tillgång till Internet ger de flesta i västvärlden möjlighet att söka kunskap och utbildning, på gott och ont kan dock denna tillgång begränsas. Tillgången till kunskap är alltså villkorad, förutom teknisk utrustningen och kunskap att använda den, krävs även ett samhälle som värdesätter yttrande- och åsiktsfrihet. Detta skapar klyftor mellan de som styr utvecklingen och de som berörs av den, tillgången till AI-teknologi och kunskap att använda den kommer troligen att få en liknande effekt. Den policy som utvecklats tyder på en vilja att behålla denna klyfta mellan västvärlden och de forna kolonierna. Incitamentet för att öka användningen av artificiell intelligens består till stor del av att öka den inhemska industrins konkurrenskraft samtidigt

som tillverkning som tidigare utlokaliserats till låglöneländer skall hämtas hem igen med hjälp av automatisering och välutbildad arbetskraft.

En aspekt av den förändrade arbetsmarknaden som inte berörs i något av dokumenten är möjligheten att även många av de nya digitala jobben går att automatisera, det är inte otroligt att en AI-lösning i framtiden kan vara den mest effektiva programmeraren eller dataanalytikern. Premissen genom hela arbetsmarknadsdiskussionen är dock att monotona och tunga arbetsuppgifter kommer att automatiseras och nya digitala arbetsuppgifter växer fram som ersättning. Den brittiska prognosen nämner en ökad arbetslöshet i Storbritannien på *endast* 10%, men att ytterligare 25% av arbetsmarknaden får räkna med stora omställningar. Kommer den förväntade produktivitetsökningen och effektiviseringen att kunna täcka upp för ökat försörjningsstöd till 10% av befolkningen, och om de nya arbetsmöjligheterna uteblir, 35%?

4.5 Algoritmer, för vem?

De analyserade policydokumenten berör inte frågan om det framväxande plattformssamhället vilket kan tyda på att författarna har svårt att tänka utanför de ramar som utgör industrialismen, utgångspunkten är att användningsområdet för artificiell intelligens främst berör automatisering och effektivisering av befintliga företag. Trots att dokumenten i svepande ordalag beskriver nya digitala sysselsättningar som vi ännu inte vet vilka de är, verkar uppfattningen vara att de traditionella företagen skall transformeras till sina digitala AI-drivna motsvarigheter. EU-kommissionen beskriver de stora vinsterna med ökad användning av artificiell intelligens inom byggsektorn och jordbruk, detta för att motivera ett underlättande för företag att använda offentliga register och utbyta lagrad personlig information som hälsodata, däremot saknas diskussionen om vilka som har störst nytta av detta. I det digitala fattighus som Eubanks målar upp argumenterar hon för att det är de sämre lottade i samhället som både får lämna ifrån sig mest personlig information och dessutom negativt särbehandlas utifrån detta (Eubanks 2017:188). Detta gör att det går att ifrågasätta om de vars data läkemedelsbolagen använder för sin forskning kommer att ha råd med behandlingarna som utvecklas, vilka kommer utifrån algoritmers avancerade gissningar nekas olycksfallsförsäkringar beroende på bostadsområde, etnisk bakgrund eller utbildning.

Mycket potentiellt lidande kan troligen undvikas genom de algoritmer vilka utvecklas för att söka igenom sociala mediaflöden och där filtrera ut användare som med hög sannolikhet kommer att drabbas av depression eller andra mentala sjukdomstillstånd. I fel händer kan denna ökade sårbarhet istället utnyttjas av företag för profit eller av totalitära stater för politiskt kontroll. Denna typ av automatiserad analys arbetar inte utifrån individnivå utan ser mönster i stora datamängder. AI-teknik bygger alltså på sannolikhet inte sanning, och i all statistik finns också en felmarginal att ta hänsyn till, men till skillnad från mänskliga beslut är det i stort sett omöjligt att förstå denna automatiserade beslutsprocess

(Andersson Schwarz, Larsson 2019:103). Detta ställer förutom höga krav på ansvar från de som utvecklar och kontrollerar dessa system även etiska regler såväl som sanktionsmöjligheter vid eventuella brister. Många yrkesgrupper så som läkare och präster, vilka i sitt arbete tar del av känsliga uppgifter omges av såväl tystnadsplikt som etiska regler. De uppgifter vilka algoritmer kan härleda utifrån dataflöden kan innehålla såväl hälsoomdömen som ideologiska och religiösa åsikter, idag finns inga klara regler för hur detta ska hanteras.

Unless they are properly regulated, in the near future we could be hired, fired, granted or denied insurance, accepted to or rejected from college, rented housing and extended or denied credit based on facts that are inferred about us (Tufekci. 2019).

Policydokumenten hänvisar till behovet av etik och hänvisar till lagar och regler, som GDPR och *the US Do Not Track Online Act of 2011*, vilka styr när och hur personuppgifter får bearbetas, men det behövs även regler för hur denna typ av härledd information får användas. När användare inte längre får möjlighet att tacka nej, utan personliga profiler skapas utifrån data utanför den personliga kontrollen behövs etiska riktlinjer för att ge enskilda bestämmanderätt över denna information (Tufekci 2019).

I den amerikanska senaten lades under våren 2019 ett förslag fram vilket adresserar algoritmer specifikt, *Algorithmic Accountability Act*, där det föreslås att större företag som använder dataset med fler än en miljon individer skall utvärdera algoritmerna utifrån bias (Paul. 2019). Detta knyter väl an till en brist hos policydokumenten, där saknas en diskussion om hur och vilka uppgifter som får användas och för vilket syfte. De handlar istället om behovet av att medborgarna skall kunna lita på objektiviteten hos algoritmerna. Det är inte tekniken i sig som beskrivs som problematisk, utan det framställs som ett problem hos användarna att ta till sig ny teknik, lita på den och lämna ut sin personliga information.

De organisationer som skapat dessa policydokument är också i högsta grad institutioner, de går att se organisatoriska strukturer där gemensamma regler upprätthålls, ofta genom någon form av bestraffning vid regelbrott. Dessa regler både skapar normer och beslutas utifrån normer, och liksom Foucaults panoptikon efterföljs dessa normer utifrån självkontroll och kontroll av andra, med tillhörande belöningar och kostnader. Utifrån detta och de förväntningar vilka finns på andras beteende formas institutionens strategier för framtiden, detta kan fortgå så länge tillräckligt många i organisationen håller fast vid normerna (Cairney 2012:74). Inom policyforskningen finns tanken på tre grader av förändring, där de två första endast är mindre justeringar utifrån befintlig policy, men där mål och problemställningar fortsatt är kvar. Vid tredje gradens förändring handlar det istället om ett paradigmskifte, där förutsättningarna drastiskt förändras (Cairney 2012:228). Frågan är om inte dessa dokument i grunden representerar de första två graderna av policyförändringar, beslutsfattarna är fortfarande fast i gemensamma regelverk och normer även när de i grunden beskriver ett paradigmskifte.

5 - Slutsatser

- Vad kan policydokument om artificiell intelligens berätta om organisationernas syn på frågor som personlig integritet, etik och moral?

Först och främst visar en sammantagen analys av dokumenten på en bred uppfattning att den kommande AI-revolutionen är både oundviklig och önskvärd. Dokumenten avspeglar en syn på artificiell intelligens liknad vid en evolutionär naturkraft, en fullständiga omdaning av både samhället och ekonomin vilken kommer att vara av minst samma magnitud som när ångmaskinen möjliggjorde industrialismen. Ställd inför risken att bli lämnad i bakvattnet av mer framgångsrika nationer finns därför en reell riska att de etiska och moraliska aspekterna sätts på undantag. Den nytta som nämns i dokumenten är oftast kopplad till effektiviseringar av företag och förvaltning, en utveckling som är beroende av utökade mängder av data, i många fall personlig sådan. Organisationerna proklamerar i dessa sammanhang, speciellt inom EU-gemenskapen, för en ökad användning av all den data som skapas inom den offentliga verksamheten, inbegripet hälsodata. Samtidigt höjs ett varnande finger för det som benämns algoritmers bias, hur artificiell intelligens uppvisar rasistiska och sexistiska tendenser. Lösningen på detta kan, tvärt emot den gängse uppfattningen, vara att använda en mindre men kontrollerad mängd med data att lära upp algoritmerna med. En större mängd ger endast mer av samma, vilket endast leder till en än större övertygelse om ett riktigt resultat, trots bias. En annan aspekt är att data som hela tiden ackumuleras gör att äldre delar av denna data reflekterar tidigare föreställningar, vilket gör att förändring av systemens bias blir fördröjd i förhållande till samhällets utveckling.

Plattformsföretag kan ses som exempel på svårigheten att förutse vad som händer när det digitala i form av algoritmer gör sitt intåg i en bransch, AirBnB som bokar övernattningar men som saknar hotell, Uber och Luft, globala taxiföretag som saknar egna bilar. En helt ny ekonomi som bygger på helt nya förutsättningar, kommer sjukvården, skolor och offentlig förvaltning att genomgå samma form av transformering? Kommer läkare att automatiseras, all utbildning att skötas online och beslutsfattande att skötas med bias-certifierad algoritmer? Går det att sätta tillbaka locket på Pandoras ask när den väl är öppnad? Detta är frågor som besvaras av dokumenten med att den personliga integriteten skall respekteras, men beslutsfattarna verkar samtidigt anse att personlig data har för högt ekonomiskt värde för att inte användas. Anonymisering föreslås för en säker hantering av känsliga uppgifter, men som forskning har visat kan saknade uppgifter härledas, den åtgärd som återstår är transparens och etiska regler som kan följas upp.

Att utveckla dessa etiska regler överlämnas i stort sett till systemutvecklarna själva att hantera, så länge som de följer de grundläggande mänskliga rättigheterna och inte uppenbart bryter mot nationella lagar. Det framstår i dokumenten som om det ligger i tekniksektorns eget intresse att kunna uppvisa gemensamma branschregler för att öka allmänhetens tilltro till produkterna. Det går även att ana en

rädsla från myndigheternas sida att om de börjar diktera strikta etiska riktlinjerna, riskerar de att marknaden inte fullt ut kan nyttja den ekonomiska potentialen hos artificiell intelligens. Det framstår därför som att de sätter sin förhoppning till att problem med bias och personlig integritet löser sig eftersom.

- Vilken effekt kan dessa policydokument förväntas ha på samhället?

En policy kan ses som ett sätt att avlasta de högsta instanserna i en organisation, regeringen, bolagsstyrelsen etc. Istället för att varje ny fråga som uppkommer längre ner i organisationen skickas upp i hierarkin för beslut, förväntas istället tjänstemän på gräsrotsnivå förhandla fram lösningar utifrån de policydirektiv som är relevanta för problemställningen. Lyckas beslutsfattarna med att skapa normer som att främst natur- och datavetenskap är nödvändiga för en framtida anställning, finns även en risk att såväl lärosäten som studievägledare och framtida studenter påverkas av detta. Artificiell intelligens framhålls som den drivande kraften för att stärka den globala ekonomin, kommer därför allt mer av forskningsanslagen i framtiden att öronmärkas för STEM-forskning? Med tanke på hur likartade dessa dokument är, både när det gäller problemformuleringar och rekommendationer, är det tydligt att detta är den norm som beslutsfattare i västvärlden arbetar utifrån. Algoritmer som resurs för att skapa ekonomiskt välstånd är det huvudsakliga fokuset, problem som kan uppstå är hur behovet av data skall tillgodoses och hur en ökad arbetslöshet skall hanteras. De etiska och juridiska problem som kan uppstå när utvecklingen går mot allt mer autonoma system, självkörande bilar, administrativa beslutssystem och sexdockor. Detta är produkter som behöver diskuteras etisk redan innan de utvecklas. De policydokument som analyserats kan därför till viss del osynliggöra de socioekonomiska problemen som finns och främst fungera som stöd för en fortsatt teknisk utveckling.

5.1 Avslutande reflektion

För att återknyta till sjätte steget i metoden WPR, vilka är förespråkarna för denna policy och vilka ifrågasätter den, vilka är vinnarna och förlorarna? Det enkla svaret är troligen att beslutsfattarna bakom policyn är både förespråkare och vinnare, men frågan är i vilken utsträckning lösningarna verkligen motsvarar problemen? Den digitala tekniken är idag så normaliserad att den sällan ifrågasätts så länge den fungerar, detta innefattar troligen även artificiell intelligens, om inte idag så inom en snar framtid. Normerna de granskade dokumenten visar på är satta utifrån marknadens behov av arbetskraft och råmaterial, digitalt kompetenta anställda och enorma mängder information om samhället. Dessa normer är bara delvis baserade på generell mänsklig etik, som lojalitet mot den egna gruppen och tjänster som gynnar den egna stammen. Här saknas dock värden som rättvisa och respekt för egendom, utifrån tanken att även personlig data går att se som personlig egendom. Algoritmerna behöver data för att uppnå sin fulla potential, men är verkligen problemet att befolkningen inte lämnar ifrån sig nog med

data? Problemet kan istället ses utifrån vilken potential som avses, dess ekonomiska eller humanitära potential. På liknande sätt går det att ifrågasätta idén om kapprustning, en företeelse som historiskt sett främst gynnar vapenhandlare. Om vi avser att nyttja algoritmernas fulla humanitära potential bör vi istället samarbeta, ett samarbete som bör föras över såväl stats- som fakultetsgränser. Slutligen frågan om framtidens arbetsliv, premissen är att en teknisk utbildning är nödvändig för att vara anställningsbar på en framtida arbetsmarknad. Borde inte kreativa utbildningar som uppmuntrar till mänsklig utveckling minst lika användbara? Om artificiell intelligens kommer att bli ett paradigmskifte, bör då inte konstruktion och underhåll av robotar också kunna automatisera?

Nick Boström och Max Tegmark kan mycket väl ha rätt i att artificiell intelligens är ett hot mot mänskligheten, inte så som det porträtteras i litteratur och på film, men hur det potentiellt kan öka både åsikts- och ekonomiska klyftor i samhället. Konflikten kommer dock att vara mellan grupper av människor— inte med robotar, lidandet kommer exklusivt att vara mänskligt. Genom att istället förändra policyarbetet mot en mer humanitär inriktning, finns förutsättningar för en framtida utvecklingen av artificiell intelligens i hela mänsklighetens tjänst. Kraftfull reglering av algoritmers användning där *ethics by design*⁷ är en av grundbultarna förhindrar inte innovation, den anpassar tekniken till användaren istället för det motsatta. Denna målsättning att genom samarbete skapa ett bättre, mer rättvist, digitalt landskap minskar troligen motståndet mot att nyttja den data som skapas. Genom ett samarbete som bedrivs öppet och med gemensamma etiska riktlinjer, kan artificiell intelligens bli ett stöd i arbetslivet, istället för att ersätta arbetskraft.

⁷ Redan på planeringsstadiet av ett system skall etiska frågor lyftas och sedan tas hänsyn till genom designprocessen.

Källhänvisningar

- Andersson Schwarz, Jonas (red) & Larsson, Stefan (red). (2019) *Plattformssamhället*. Stockholm: Fores
- Bacchi, Carol. (2009) *Analysing Policy: What's the Problem Represented to Be?* French Forest: Pearson Australia
- Bacchi, Carol. (2012) Why Study Problematizations? Making Politics Visible. i *Open Journal of Political Science* 2(1) 1-8. DOI: 10.4236/ojps.2012.21001
- Bergström, Göran och Boréus, Kristina. (2005) *Textens mening och makt*. (2:a utg.) Lund: Studentlitteratur
- Boréus, Kristina. (2015) Diskursanalys. i Ahrne, Göran (red). Svensson, Peter (red). *Handbok i kvalitativa metoder*. (2:a utg.) Stockholm: Liber 176-190
- boyd, danah och Crawford, Kate. (2017) Critical Questions for Big Data i *Information, Communication & Society* 15(5) 662-679 DOI: 10.1080/1369118X.2012.678878
- Börjesson, Mats. (2003) *Diskurser och konstruktioner*. Lund: Studentlitteratur
- Cairney, Paul. (2012) *Understanding Public Policy*. Houndsmills: Palgrave MacMillan
- Colebatch, H. K. (2009) *Policy*. (3:e utg.) Maidenhead: Open University Press
- Collste, Göran. (2002) *Inledning till etiken*. (2:a utg.) Lund: Studentlitteratur
- Eubanks, Virginia. (2017) *Automating Inequality*. New York: St. Martin's press
- Foucault, Michel. (1991) *Discipline and punish : the birth of the prison*. Harmondsworth: Penguin
- Gelin, Martin. (2018) Kina har blivit bäst i världen på att censurera internet. i *Dagens Nyheter*. 2018-03-27 Del 2, sidan 8
- Goertzel, Ben och Pennachin, Cassio. (2007) *Artificial General Intelligence*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg
- Hanell, Fredrik. (2018) What is the 'problem' that digital competence in Swedish teacher education is meant to solve? i *Nordic Journal of digital literacy*. 13(3) 137-151
- Isaacson, Walter. (2014) *Innovatörerna*. Stockholm: Bonniers Förlag
- Kranzberg, Melvin. (1986) Kranzbergs laws. i *Technology and Culture* 27(3) 544-560 DOI: 10.2307/3105385
- Lenas, Sverker. (2017) Max Tegmark rädd att skräckscenariot med artificiell intelligens ska bli verklighet. i *Dagens Nyheter*. [e-resurs] 2017-12-05 <https://www.dn.se/kultur-noje/max-tegmark-radd-att-skrackscenariot-med-artificiell-intelligens-ska-bli-verklighet/> Läst 2019-04-10
- Liedman, Sven-Eric. (1997) *I skuggan av framtiden*. Stockholm: Bonnier Pocket
- Lovén, Svante. (1999) Den tänkande maskinen. *Tidskrift för litteraturvetenskap*. 29(3-4) 19-38
- Machin, David och Mayr, Andrea. (2012) *How To Do Critical Discourse Analysis*. London: Sage

- Michie, Donald. (2008) Alan Turing's Mind Machines. i Husbands, Philip (red). Holland, Owen (red) och Wheeler, Michael (red). *The Mechanical Mind in History*. Cambridge, Mass.: MIT Press 61-74
- Miller, Vincent. (2011) *Understanding Digital Culture*. London: Sage Publishing
- Molander, Joakim. (2003) *Vetenskapsteoretiska grunder*. Lund: Studentlitteratur
- Nepolitan, Richard E. och Jiang, Xia. (2018) *Artificial Intelligence*. (2:a utg.) New-York: Chapman and Hall
- Noble, Umoja Safiya. (2018). *Algorithms of Oppression*. New York: New York University Press
- Paul, Kari. (2019) Disastrous' lack of diversity in AI industry perpetuates bias, study finds. i *The Guardian.com* [e-resurs] <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study> Läst 2019-04-22
- Pripp, Oscar. (2011) Reflektion och etik. i Kaijser, Lars (red). Öhlander, Magnus (red). *Etnologiskt fältarbete* (2:a utg.). Lund: Studentlitteratur, 65-84
- Scott Curry, Oliver. Austin Mullins, Daniel och Whitehouse, Harvey. (2019) Is It Good to Cooperate? i *Current Anthropology* 60(1) 47-69 DOI: 10.1086/701478
- Strannegård, Lars (red) och Eriksson-Zetterquist, Ulla (red). (2011) *Organisering*. Malmö: Liber
- SVT. (2018) *Detta har hänt: Facebook och Cambridge Analytica* [e-resurs] <https://www.svt.se/nyheter/utrikes/detta-har-hant-facebook-och-cambridge-analytica> Läst 2019-04-09
- Tufekci, Zeynep. (2019) Think You're Discreet Online? Think Again. i *New York Times* [e-resurs] <https://www.nytimes.com/2019/04/21/opinion/computational-inference.html> Läst 2019-04-21
- Van Dijck, José. Poell, Thomas. De Wall, Martijn. (2018) *The Platform Society*. New York: Oxford University Press

Policydokument

- European Commission. (2018) *Artificial Intelligence for Europe. COM(2018)237 final* [e-resurs] <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>
- Executive Office of the President. (2016) *Artificial Intelligence, Automation, and the Economy* [e-resurs] <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF> Läst 2019-04-04
- Government office for science. (2016) *Artificial Intelligence: opportunities and implications for the future of decision making* [e-resurs] <https://www.gov.uk/government/publications/artificial-intelligence-an-overview-for-policy-makers>, Läst 2019-04-01
- IBM. (2017) *IBM Cognitive Principles* [e-resurs] https://www.ibm.com/ibm/responsibility/ibm_policies.html, Läst 2019-04-07
- Intel. (2017) *Artificial Intelligence; The Public Policy Opportunity*. [e-resurs] <https://www.intel.ai/ai/wp-content/uploads/sites/69/Intel-AI-Public-Policy-WP-2017.pdf> Läst 2019-04-07