ROLF LIDSKOG

Samhället utmanat?[1]

Artificiell intelligens och sociologisk kunskap

Society challenged? Artificial intelligence and sociological knowledge
Artificial intelligence (AI) is a growing social phenomenon, both as a technical infrastructure and as a sociotechnical imaginary. While there is ample sociological research on AI's implications for different societal sectors, there has been limited sociological discussion on AI itself. This article engages with the current debate on superintelligence – a future AI that transcends human intelligence – to show that the development of AI concerns central sociological issues, and that sociological thinking contributes to and challenges prevailing concepts of AI.

Keywords: artificial intelligence, deep learning, machine learning, social intelligence, superintelligence

ARTIFICIELL INTELLIGENS (AI) är ett område som fått allt större uppmärksamhet. Det finns stora förhoppningar knutna till dess utveckling och stat och näringsliv gör mycket stora satsningar.[2] Samtidigt finns en medvetenhet om att AI kan användas för destruktiva ändamål och därför behöver forskning och utveckling på detta område regleras. Dessa farhågor utgör även argument för att Europeiska unionen och Sverige bör satsa kraftfullt så att demokratiska värderingar styr denna teknikutveckling.[3]

1 Jag vill tacka Monika Berg och Sverre Wide (Örebro universitet) samt tre granskare för konstruktiva kommentarer på tidigare versioner av denna artikel.

2 Ett aktuellt svenskt exempel är Wallenbergstiftelsernas pågående satsning på grundforskning om AI och kvantteknologi (totalt drygt 4 miljarder kronor). I satsningen ingår även 660 miljoner för samhällsvetenskaplig och humanistisk forskning om utmaningar och effekter av AI och autonoma system.

3 I TV-programmet "De smarta maskinernas tid" (sänt i *Vetandets värld* på SVT den 27 augusti 2018), säger Sveriges näringsminister Mikael Damberg att "Ny teknik innebär inte med automatik att det är gott eller illa, det måste fyllas med värderingar och moral. Det kan vi i Europa och Sverige bidra med". Han oroas också över AI-forskningen i Kina och säger att "Kina har ett helt annat samhällssystem och syn på mänskliga rättigheter än Sverige. Vi måste därför fylla AI med moraliska värderingar, och sätta gränser". I samma program intervjuas AI-forskaren Max Tegmark som betonar att "det krävs krafttag av våra politiker, inte bara för att vi [inte] ska halka efter ekonomiskt utan också för att våra europeiska värderingar ska vara med att styra den här tekniken".

https://doi.org/10.37062/sf.57.19591
Sociologisk Forskning, årgång 57, nr 2, sid 99–120.

Men AI är inte enbart en socioteknisk föreställning innehållande förväntningar och farhågor. Den är även en befintlig teknik och infrastruktur. Teknik- och vetenskapssociologin har sedan lång tid tillbaka betonat att samhället inte bara är befolkat av människor och organisationer utan även av tekniska artefakter (Collins & Pinch 1998; Latour 2005). Ska vi förstå samhällets utformning och det sociala livet måste vi studera de tekniska system som är med och konstituerar det moderna samhället. Att studera hur AI är del i formandet och omformandet av samhället är därför en viktig – och allt viktigare – sociologisk uppgift.

Det finns en ansenlig mängd forskning om digitaliseringens konsekvenser, inte minst för arbetslivet (Braun, Zweck & Holtmannspötter 2016; Broadbent 2016; Frey & Osborne 2017; McClure 2018) och samhällslivet (Broadbent 2016; Hintz, Dencik & Wahl-Jorgensen 2019; Lupton 2014; Makridakis 2017). Det finns däremot få sociologiska studier som diskuterar AI mer generellt. Det finns äldre sociologiska bidrag om AI och maskininlärning som socialt fenomen (till exempel Bainbridge, Brent, Carley m.fl. 1994; Brent 1988; Carley 1996; Collins 1990; Schwartz 1989; Woolgar 1985). På svenska är troligtvis Johan Asplunds (2002) socialpsykologiska kritik av AI den mest kända sociologiska analysen medan Peter Kåhres (2009) systemteoretiska perspektiv på AI är mer okänt. Två nyligen publicerade internationella forskningsöversikter finner att merparten av de samhällsvetenskapliga studierna om AI kännetecknas av att antingen ha ett begränsat fokus eller bygga på teknisk determinism (Boyd & Holton 2018; Mlynař, Alavi, Verma m.fl. 2018). Dessutom finner de att bidragen ofta är ateoretiska och ahistoriska i synen på relationen mellan teknik och samhälle.

Bakgrunden till denna artikel är den allt centralare roll AI har i samhället, det är ett sociotekniskt system som blir alltmer invävt i samhällsverksamheter och det sociala livet. Syftet är att bidra till utvecklingen av detta forskningsfält genom att dels visa att utvecklingen av och forskningen om AI berör centrala sociologiska frågor, dels visa att sociologiskt tänkande bidrar till och utmanar vissa föreställningar om AI. Artikeln fokuserar dock inte alla de förväntningar och farhågor som är knutna till AI:s utveckling. I stället tar den sin utgångspunkt i en av de mest långtgående föreställningarna: att AI kan utveckla en intelligens som är överlägsen människans (vad AI-forskare benämner ”superintelligens”). Denna föreställning väcker grundläggande sociologiska frågor om och i vad mån tekniska system kan tänka (intelligens) och handla (agens) – och vad som i så fall menas med dessa begrepp.

Det är dock ingen enkel uppgift att definiera vad AI är eftersom det såväl vetenskapligt som i den offentliga debatten tilldelas olika innebörder. Inom viss forskning används akronymen STAARA – *Smart Technologies, Artificial intelligence, Automation, Robotics, Algorithms* – för att fånga att dessa tekniker tillsammans utgör en teknikfamilj (se till exempel Brougham & Haar 2017). Den mest elementära definitionen av AI är att den avser icke-biologisk intelligens, men denna definition väcker frågor om vad intelligens är, vilket är en omdiskuterad fråga inom såväl AI-forskning som hjärnforskning, psykologi och sociologi. Jag väljer dock att i denna artikel utgå från denna elementära definition av AI – en intelligens som har sin grund i en materiell artefakt av icke-biologisk karaktär.

En lika svår uppgift som att definiera AI är att avgränsa den som studieobjekt. Forskningslitteraturen har olika disciplinrelaterade utgångspunkter (från fysik till antropologi), berör en mängd olika områden samt har olika målgrupper (från den egna disciplinen till den breda allmänheten). Min beskrivning av AI-området är baserat på några grundläggande böcker om AI (Goodfellow, Bengio & Courville 2016; Kelleher 2019; Sejnowski 2018) men framför allt på två böcker som fått stort genomslag i den vetenskapliga och den offentliga debatten.[4] Den ena är *Superintelligens. Vägar, faror, strategier* (2017[2014]) och är skriven av Nick Bostrom, professor i filosofi i Oxford. Den andra är *Liv 3.0. Att vara människa i den artificiella intelligensens tid* (2017[2017]) och är skriven av Max Tegmark, professor i fysik vid MIT. Båda dessa böcker handlar om AI som samhällsutmaning och är skrivna utifrån ett filosofiskt-etiskt perspektiv (Bostrom) respektive ett reflekterat AI-perspektiv (Tegmark). I dessa böcker och i vetenskapliga artiklar, men även i medier och offentliga diskussioner, argumenterar de för att samhället bör göra stora satsningar på att såväl utveckla AI som att reglera denna utveckling.[5] De anser att frågan om AI, inte minst superintelligens, kan vara den största utmaning som mänskligheten någonsin ställts inför och att AI därför bör vara vår tids viktigaste samtalsämne (Bostrom 2017:10; Tegmark 2017:29).

Utöver detta material har en litteratursökning genomförts med syfte att finna vetenskapliga artiklar som diskuterar AI ur ett sociologiskt perspektiv.[6] Syftet med litteratursökningen har varit explorativt: att skapa en översiktlig bild av vad sociologer har att säga om AI som samhällsfenomen.

Artikeln disponeras i fyra avsnitt med denna inledning inräknad. I nästa avsnitt beskrivs kortfattat maskininlärning och föreställningen om AI och superintelligens. Fokus på superintelligens betingas dels av att flera AI-forskare ser den som den viktigaste samhällsutmaningen, dels av att denna diskussion väcker många generella frågor om AI. Efter att ha låtit AI-forskare själva komma till tals vänds i det tredje avsnittet intresset till sociologiska bedömningar av AI. Diskussionen är här inriktad på frågor, förväntningar och farhågor som superintelligens väcker. I det avslutande avsnittet betonas att ett alltför stort fokus på AI:s förväntade utveckling kan innebära att frågor om samhällets användning av och anpassning till AI inte får tillräcklig uppmärksamhet.

4 Bostroms och Tegmarks böcker placerades på *New York Times* bestsellerlista och Tegmarks bok på *Times* lista över årets bästa böcker. Båda dessa böcker är översatta till en mängd språk, däribland svenska.

5 De arbetar med att föra upp AI-frågan på politiska och offentliga dagordningar och har skapat institut kring frågor om AI och framtiden. I Sverige ses de kanske som de främsta svenska AI-forskarna och de har båda varit sommarpratare i Sveriges Radio (Tegmark år 2017 och Bostrom år 2019).

6 Litteratursökningen gjordes ursprungligen för perioden 2008–2018 i databaserna Sociological Abstracts och Social Science Premium Collection, men har uppdaterats med en ny sökning för 2019 (genomförd i januari 2020). Därutöver har osystematiska sökningar gjorts i Web of Science och Google Scholar som resulterade i ytterligare träffar. Totalt har 47 artiklar och tio böcker analyserats.

Artificiell intelligens

I dagens debatt om tekniska innovationer och social förändring finns två dominerande positioner (Boyd & Holton 2018). Den första är att den teknikutveckling vi ser i dag inte innebär en grundläggande samhällstransformation. AI är givetvis betydelsefull för samhällets utveckling, precis som många tidigare tekniska innovationer varit det. IT-revolutionen har redan genomförts och även om den har förändrat och förändrar arbetsmarknad och vardagspraktiker har den inte i grunden förändrat samhället. Den andra positionen hävdar det motsatta: att vi nu står för en fundamental samhällstransformation. Klaus Schwab (2017), professor i företagsekonomi och grundare av World Economic Forum, hävdar att samhället har haft tre industriella revolutioner, skapade av ångkraft, elektricitet och digitalisering, och nu står inför en fjärde som är mer omfattande och komplex än tidigare: AI och robotik. Genom sammanflätningar av teknik, naturvetenskap och biologi kommer vi inte bara skapa nya produkter och sätt att producera utan även på djupet förändra samhället. Många AI-forskare omfattar denna position, att AI kommer att ha revolutionerande och kanske till och med oöverblickbara samhällskonsekvenser.

Djup maskininlärning

Vad gäller vissa kognitiva uppgifter, inte minst beräkningskraft och hastighet, har datorer sedan en lång tid tillbaka överträffat människan.[7] Det handlar här om det som AI-forskare kallar *snäv intelligens*, en intelligens utvecklad för ett speciellt syfte. Ett exempel är här schackdatorer som sedan lång tid tillbaka slår de bästa schackspelarna i världen. Men dessa schackdatorer kan inte utföra andra arbetsuppgifter, inte ens spela andra spel. Människan har däremot *generell intelligens* i bemärkelsen att hon inte är (genetiskt) förprogrammerad för att enbart utveckla kognitiva färdigheter inom ett avgränsat område. I detta sammanhang förs oftast fram att datorer och artificiell intelligens främst handlar om *ytlig maskininlärning.* Människan kan genom instruktioner och programmering lära en maskin olika saker, och den kan till exempel upptäcka mönster i stora mängder data (till exempel datautvinning ur *big data*). Men ytlig maskininlärning innebär att datorn inte kan ta egna initiativ till lärande utanför det område som människan ursprungligen instruerat den till att agera inom.

Djup maskininlärning är ett kvalitativt annorlunda lärande där maskiner självständigt utvecklar sin egen intelligens (Goodfellow, Bengio & Courville 2016; Sejnowski 2018). Denna typ av lärande har tidigare setts som unikt mänsklig men AI-forskningen finner det allt svårare att dra en tydlig gräns mellan mänsklig och artificiell intelligens. Genom att basera ett AI-system i artificiella neuronnätverk kan det gradvis lära sig nya saker och även utveckla sin egen arkitektur (konstruera nya algoritmer och be-

7 Vartannat år har antalet transistorer på integrerade kretsar fördubblats (Bostrom 2017:51) och mängden maskinella beräkningar har halverats i pris (Tegmark 2017:89). Polson och Scott (2018) gör analogin att om bilars toppfart hade ökat i samma takt som datorerna skulle en bil med toppfarten 180 kilometer i timmen år 1951 i dag ha toppfarten av åtta miljoner gånger ljusets hastighet.

räkningsstrukturer) som i sin tur förbättrar dess kognitiva och motoriska prestationer. Djup maskininlärning innebär att AI till synes självständigt kan lära sig en mängd olika färdigheter. Om den utvecklas till att få förmågan att självständigt (utan mänsklig insats) och ständigt förbättra sin intelligens (vad AI-forskare benämner ”rekursiv självförbättring”) kan AI kontinuerligt utveckla en allt bättre (effektivare, smartare, kraftfullare) version av sig själv.

Som fallet är för många tidigare tekniska innovationer utvecklas det även kring AI en mängd föreställningar om dess potential. Dessa föreställningar är performativa i bemärkelse att de påverkar såväl riktningen för som regleringen av AI:s utveckling. Jag vill här fästa uppmärksamhet på en av de mest långtgående föreställningarna: att AI inte bara kan utveckla en intelligens på mänsklig nivå (stark AI) utan en intelligens som vida överstiger människans (superintelligens).

Superintelligens

Det finns i dag en mängd applikationprogram och teknologiska uppvisningar (demonstrering av en ny teknik i syfte att visa dess potentialitet) baserade i djup maskininlärning där AI-systemet tillägnar sig nya kompetenser och färdigheter utan att vara förprogrammerade för att lära sig dessa. Dessa AI-system benämns ofta autonoma och intelligenta. För de flesta sammanhang anses detta vara eftersträvansvärt, men det finns även risker förknippade med denna utveckling. Därför arbetar många AI-forskare med att skapa robusta AI-system: system som dels utför det de är avsedda att göra och dels gör det utan att fela.[8] En fråga som fått allt större uppmärksamhet är dock av annan karaktär. Vad händer om AI utvecklar en intelligens som överskrider människans? AI-forskare talar här om en ”intelligensexplosion” eller ”singularitet” (från engelskans *singleton*). Oron är här att när AI kan förbättra sig själv kommer dess intelligensutveckling att gå allt snabbare och leda till en generell intelligens på betydligt högre nivå än människans (”superintelligens”). Om denna intelligens inte enbart kan utveckla sin mjukvara utan även sin hårdvara innebär det även att den kan bli oberoende av mänskliga insatser för sin aktivitet och reproduktion.

I sin bok *Liv 3.0* tar Tegmark sin utgångspunkt i just denna tänkbara utveckling. Som titeln visar delar han in livet i tre stadier: Ett *biologiskt*, som inleddes med livets uppkomst för 4 miljarder år sedan (Liv 1.0), ett *kulturellt*, med människans uppkomst för ungefär 100 000 år sedan (Liv 2.0) och ett *tekniskt*, som kan uppkomma under det kommande århundrandet (Liv 3.0) (Tegmark 2017:32–39). Liv 1.0 är utvecklat genom evolution, Liv 2.0 är en kombination av evolution och självskapelse medan Liv 3.0 enbart är självskapat, det vill säga det finns inga biologiska processer involverade i uppkomsten, vidmakthållandet och utvecklingen av denna typ av liv. Liv 3.0 innebär att de absoluta gränser som finns för liv inte längre är biologiska (eller sociala) utan

8 Tegmark (2017:124–139) beskriver fyra huvudområden inom teknisk AI-säkerhetsforskning: verifiering (säkerställa att ett program/system uppfyller alla förväntade krav), validering (att programmet baseras på antaganden som är giltiga), säkerhet (mot såväl externa attacker som interna buggar) och kontroll (något eller någon som övervakar ett system och vid behov ändrar dess beteende).

enbart fysiska.[9] AI kan överskrida de biologiska gränser som finns för människan. Det är alltså en enorm potential som här tillmäts AI och Tegmark ser den pågående utvecklingen, inte minst utvecklingen av artificiella neuronnätverk, som ett led mot denna utveckling. Tegmark (ibid.:426) gör en målande beskrivning av vad AI:s utveckling kan innebära:

> Först hade vi människor upptäckt hur vi kunde kopiera vissa naturliga processer med maskiner, skapa våra egna vindar och blixtar och vår egen mekaniska hästkraft. Gradvis började vi inse att våra kroppar också var maskiner. Sedan började upptäckten av nervceller sudda ut gränsen mellan kropp och medvetande. Därefter började vi bygga maskiner som kunde utklassa inte bara våra muskler, utan även vårt medvetande. Så parallellt med att upptäcka vad vi är håller vi alltså oundvikligen på att göra oss själva förlegade? Det vore djupt tragiskt.

Den stora faran är att denna superintelligens utvecklar nya mål och verksamheter som inte ligger i linje med dess ursprungliga (av människan skapade) syfte och där den dessutom kan utveckla egna instrumentella mål som gör den ytterst svårstoppad, till exempel mål som självbevarande och bevarande av sitt slutmål (Bostrom 2017:167–182). Kan den utveckla sin egen mjuk- och hårdvara har den en autonomi som gör den oerhört kraftfull och robust mot yttre påverkan och kan reproducera sig själv i allt intelligentare och robustare form. Hotet är således inte att ett AI-system söker världsherravälde utan att den utvecklar mål som inte står i samklang med människans och att den dessutom har förmåga att uppnå dessa mål.

Frågan om en framtida intelligensexplosion och uppkomsten av superintelligens väcker dock inte bara farhågor, utan även förväntningar och förhoppningar. Bostrom ser en extrem potential i AI, till exempel att människan inte bara kan leva betydligt längre utan även vara vid god hälsa och utveckla nya fysiska och kognitiva förmågor (så kallad mänsklig förbättring, *human enhancement* (Bostrom & Sandberg 2017)). Tegmark diskuterar hur AI kan lösa flera av dagens globala överlevnadsfrågor, till exempel möjliggöra en kolonialisering av rymden vilket ger tillgång till i närmast oändliga naturresurser.[10]

9 Utgångspunkten för Tegmarks resonemang är att liv, medvetande och intelligens inte har någon nödvändig relation till biologi. *Liv* definierar Tegmark (2017:32) inklusivt, som en process som kan bibehålla sin komplexitet och kopiera sig själv, och det som kopieras är materiens struktur (information). På art- och populationsnivå finns här en stor överensstämmelse med en biologisk definition av liv: förmåga till reproduktion och evolution (överfört till AI: förmåga till reproduktion och rekursiv självförbättring). På individnivå har dock biologin en betydligt mer krävande definition, till exempel förekomst av celler och förmåga till ämnesomsättning (Taylor, Simon, Dickey m.fl. 2017). Den begreppsinnebörd som Tegmark ger liv – reproduktion av information – kräver däremot ingen biologisk grund.

10 Tegmark (2017:293) anger att om vi nyttjar all materia i vårt solsystem (vad han kallar ”kosmiskt kapital”) har vi cirka en miljon gånger mer materia att nyttja än vad som finns på jorden. Att utvinna naturresurser från andra planeter eller åtminstone asteroider är inte bara något som AI-forskare utan även kommersiella och politiska intressen ser som möjligt. Sedan 2015 finns i USA en lag om äganderätten vid utvinning av resurser i rymden (Perdue 2017). →

Föreställningen om en intelligensexplosion kan framstå som naiv och ha karaktären av science fiction snarare än kvalificerad analys. Men Bostrom gör en systematisk genomgång av intelligensexplosionens kinetik och visar att om en intelligensexplosion sker står mänskligheten inför såväl oanade möjligheter som extrema risker. Denna utveckling är inte deterministisk, resan mot superintelligens är till en början helt beroende av mänskliga initiativ och beslut.[11] Men om denna utveckling sker är människan inte längre den intelligentaste varelsen på jorden och därmed inte heller den mäktigaste. Målet är därför att skapa en robust och välvillig superintelligens, där robust avser att den inte fallerar och välvillig att dess mål står i samklang med människans. AI-forskare betonar att dessa frågor redan i dag måste styra forskningen och utvecklingen av AI.[12] Att problemet anses vara ytterst allvarligt syns inte minst i hur Bostrom avslutar sin genomgång av superintelligensens möjligheter och faror (Bostrom 2017:399–400):

> Superintelligens är en utmaning som vi inte är redo för idag, och inte kommer att vara redo för på länge än. […] I denna situation är varje känsla av förtjust upphetsning malplacerad. Bestörtning och rädsla vore mer adekvat; men den lämpligaste attityden är kanske en sammanbiten beslutsamhet att vara så kompetenta som vi kan […] Den utmaning vi står inför är dessutom delvis att hålla fast vid vår mänsklighet: att stå stadigt på jorden och behålla vårt sunda förnuft och en godlynt anständighet även inför detta ytterst onaturliga och omänskliga problem. Vi måste använda hela vår mänskliga uppfinningsrikedom för att försöka lösa det.

→ Nasa förbereder nu en första expedition år 2022 till en asteroid främst bestående av metall (Nasa 2020). 2010-talets förväntan på en ny ”guldrush” till planeter har dock mattats av eftersom initiala kostnader och tekniska hinder hittills visat sig vara stora (Abrahamian 2019).

11 Bostrom (2017:43–87) beskriver fyra möjliga teknologiska vägar till superintelligens: hjärnemulering, biologisk kognition, sammankoppling mellan hjärna och dator (till exempel genom implantat) samt ett nätbaserad kognitivt system som kopplar samman individuella mänskliga medvetanden med datorer. Han beskriver även tre olika former av superintelligens: Snabb: intelligens på samma nivå som människans, men extremt snabbare (till exempel kapaciteten att tillgodogöra sig innehållet i en bok på tio sekunder); kollektiv: intelligens på samma nivå som människan men genom sammankoppling vida överlägsen mänsklig intelligens; kvalitativ: inte nödvändigtvis snabbare än mänsklig intelligens, men kvalitativt smartare (jämför skillnaden mellan människans och djurs intelligens).

12 Bostrom (2017:323–350) föreslår två vägar för att skapa en välvillig AI: kontroll av vad en AI kan göra (”kapacitetskontroll”) respektive vad den vill göra (”motivationsurval”). Han finner att kapacitetskontroll i bästa fall kan vara en tillfällig och kompletterande åtgärd men att långsiktigt krävs att AI-utvecklare kan inplantera någon typ av gynnsamma värden som AI eftersträvar. Denna fråga är extremt komplex, vilket Bostrom själv betonar. Det är vanskligt att i dag veta vilka värden eller mål som är önskvärda för framtida generationer och det är en stor risk att historiskt kontingenta uppfattningar om ett gott liv präglar de värden man laddar en AI med. Bostrom ser dock en lösning genom att infoga en ”indirekt normativitet”, det vill säga att ladda motivationssystemet hos en superintelligens med ett abstrakt önskvärt tillstånd som AI sedan kan finna konkreta normer för som leder till detta tillstånd. Även om detta visar sig möjligt är dock ett lika stort problem att översätta dessa värden eller mål till ett programmeringsspråk utan att viktigt innehåll försvinner.

Sociologisk bedömning

Att ta del av AI:s historiska utveckling och förhoppningar och farhågor knutna till dess framtida utveckling väcker filosofiska frågor om vad liv, medvetande och intelligens är, epistemologiska frågor om vad som utgör kunskap, politiska frågor om behovet av och villkoren för teknikreglering, etiska frågor om i vilken riktning samhället bör utvecklas och hur vi bör hantera den allt starkare sammanflätningen av maskin och människa. Men här väcks även en mängd sociologiska frågor eftersom AI till stor del handlar om gränssnittet samhälle–människa–maskin. I det följande fäster jag uppmärksamhet på fyra centrala frågor. Först diskuteras frågan om AI kan utveckla en social intelligens följt av frågan om AI kan utveckla agens. De andra två frågorna vänder intresset från AI:s eventuella utveckling till AI:s nuvarande påverkan på samhället. Det handlar om frågan om hur samhället påverkas av AI och faran med att sätta för stark tilltro till AI:s kapacitet samt frågan om vilka som styr och vad som driver dagens AI-utveckling.

Kan AI utveckla mänsklig intelligens?

Frågan om AI kan utveckla en intelligens är något som humanister och samhällsvetare diskuterat under en lång tid. Mest känd är troligtvis Hubert Dreyfus (1972, 1992) fenomenologiskt och Heidegger-inspirerade kritik av AI samt John Searles (1980) kritik av AI där han hävdar att för att klassificera ett beteende som tänkande krävs förståelse och intentionalitet. Den svenske sociologen Johan Asplund (2002) för fram en liknande kritik där hans utgångspunkt är att intelligens är ett intermentalt och kommunikativt attribut mellan människor och följaktligen kan inte AI vara intelligent och därmed inte heller ha agens (eftersom en handling innebär att man vet varför man utförde den).

Den sociolog som mest omfattande och ambitiöst diskuterat AI är den brittiske vetenskapssociologen Harry Collins. I sin bok *Artificial experts. Social knowledge and intelligent machines* (1990) hävdar Collins att AI inte kan utveckla vad han benämner *social intelligens.* För strikt regelstyrda verksamheter kan AI utveckla en intelligens som motsvarar eller överträffar människans. Men den mänskliga kunskapen har till stor del en annan karaktär: den är socialt präglad, kontextuell och ofta implicit och oartikulerad. Medan en regelstyrd verksamhet kan tillägnas genom instruktion kan den sociala kunskapen endast tillägnas genom socialisation och socialt samspel i en avgränsad gemenskap.[13] Denna kunskap är förreflexiv och oartikulerad, där en person omedvetet och gradvis tillägnar sig en förståelse över vilka sociala normer och kulturella koder som gäller för ett visst sammanhang, vilken kunskap som tillmäts värde och vilken kompetens som är relevant i detta sammanhang.[14] Det handlar om

13 Harry Collins refererar till den wittgensteinska ståndpunkten att för att kunna följa en regel på ett adekvat sätt krävs kännedom om i vilken kontext den ska följas.

14 För att tydliggöra sin poäng illustrerar Collins (1990:6–7) med ett fiktivt exempel: en brittisk spion flyttar till Semipalatinskt (stad i nuvarande Kazakstan) och utger sig för att vara uppvuxen där. Trots omfattande förberedelser där spionen har tillägnat sig all nedtecknad kunskap som finns tillgänglig om denna ort och dess befolkning är det inte tillräckligt. När lokalbefolkningen →

en intelligens som särskiljer sig från regelstyrd och algoritmbaserad intelligens. Harry Collins omformulerar därmed frågan om intelligens till frågan om kunskapens natur och vad människan inte kan artikulera.

AI ha dock utvecklats enormt sedan Collins bok publicerades för tre decennier sedan och han har nyligen utkommit med boken *Artifictional intelligence. Against humanity's surrender to computers* (2018) där han diskuterar i vad mån djup maskininlärning innebär att gränsen mellan maskinintelligens och mänsklig intelligens nu kan överskridas. I sin diskussion om vad intelligens är tar han i denna bok inte kunskapen utan språket som utgångspunkt. Människan besitter vad Collins benämner ett *naturligt språk*, ett språk som är otydligt, fragmenterat, kontextkänsligt, kreativt och regelbrytande. Ändå förstår vi varandra och det beror på att språket inte handlar om regelföljande utan om kontextuell förståelse som gör att vi kan skapa en rimlig och trovärdig mening om det vi ser och hör. Att tillägna sig ett språk innebär att man erhåller denna kontextkänslighet där ett yttrande eller en handling tolkas utifrån sitt sammanhang. Beroende på sammanhang tolkas ett budskap bokstavligt eller ironiskt och en svordom nedsättande eller uppskattande. Att behärska ett språk handlar därför inte om att uttrycka sig grammatiskt korrekt utan om att tillägna sig en tolkningsförmåga: att kunna förmedla det man vill ha sagt och förstå vad andra menar med det de säger. Collins (2018:13) anser därför att till skillnad från många av dagens tester och indikatorer på att en dator är intelligent (som till exempel Turingtestet eller Searles kinesiska rum) är en rimligare test att undersöka i vad mån en dator kan reparera (kontextkänsligt omtolka) en människas språkliga misstag.

Vad som gör denna uppgift ytterst komplicerad är att människan ständigt, oväntat och innovativt bryter mot språkliga prejudikat. Det innebär att det inte är tillräckligt att ha kunskap om ett språks historiska eller nutida användning för att kunna avgöra om ett språkligt regelbrytande är legitimt eller inte (legitimt i bemärkelse att andra förstår innebörden i vad en person söker förmedla). Djupinlärningens avancerade mönsterigenkänning är därför inte tillräckligt utan det krävs att uttolkaren är socialt inbäddad i ett språktalande samhälle.

Det finns dock sociologer som hävdar att det inte finns något principiellt hinder för AI att utveckla en social intelligens. För den amerikanske sociologen Randall Collins (2008:169–185) handlar social intelligens framför allt om att kunna socialt samspela vilket kräver att man tillägnar sig en kulturell kompetens som gör det möjligt att förstå kulturspecifika (kontextuella) uttryck, outtalade budskap och kroppslig (icke-verbal) kommunikation. Dessutom krävs att man förstår samtalets sociala karaktär (inte minst att kompetent nyttja turtagning, injustering och tajmning). Han anser att AI kan utveckla denna samtalskompetens när den kan synkronisera till en samtalspartners rytm och anknyta till dess intresse (vad han benämner gruppspecifikt kulturellt kapital) vilket i sin tur skapar positiv emotionell energi som gör att man vill

→ möter spionen kommer de direkt att upptäcka att hen inte är uppvuxen på denna plats. Detta eftersom hen i sitt sätt att vara och samspela saknar viktig social kunskap, kunskap som inte är artikulerad och som endast kan erhållas genom att leva på denna ort och samspela med dess befolkning.

fortsätta samtala. Randall Collins verkar anse att emotionell energi kan skapas utan att AI har ett eget känsloliv; det är tillräckligt att AI kan avläsa och tolka emotionella signaler och denna förmåga kan den tillägna sig genom att analysera samtal (tidigare, pågående och föreställda) och lära sig interaktionsritualer. En förutsättning för denna utveckling är att AI blir kroppslig, det vill säga att den inte bara analyserar kroppslig interaktion och kommunikation utan även själv deltar i dessa samspel. Detta är även i linje med vad Harry Collins (2018:67–73) betonar; en förutsättning för att utveckla social intelligens är att man är inbäddad i ett socialt sammanhang, vilket i sin tur kräver kroppsligt deltagande.

Randall Collins (2008) betoning av den sociala intelligensens kommunikativa karaktär är poängfylld. När AI:s kommunikativa förmåga framhålls handlar det oftast om kroppslös och röststyrd informationshantering där AI förstår och besvarar frågor inom ett välavgränsat område. Denna typ av samtal bygger på en begränsad och asymmetrisk kommunikation där människan anpassat sig till speciella kommunikativa villkor. Som Garfinkel (1984) och Goffman (1970) visat upprätthåller vi ett samtal genom att hela tiden välvilligt tolka, neutralisera och reparera det. Vi gör det oftast omedvetet genom att nyttja den kulturella, sociala och språkliga kompetens vi erhållit genom vår socialisation. Därför reflekterar vi sällan över att vi ständigt kompenserar en dators kommunikativa brister, till exempel digitala assistenters bristande förmåga att förstå vad vi säger.[15] I vissa fall blir det uppenbart för oss och vi accepterar det (räddar interaktionsordningen) just genom att vi vet att det är en digital assistent och inte en människa vi samtalar med. Människan skapar därmed en AI-illusion: att AI är mer språkligt kompetent och socialt intelligent än vad den är. Det handlar om en svag AI-socialitet: AI kan socialt samtala och samspela givet att människan anpassar sig till AI:s begränsade kommunikativa och interaktiva repertoar (Rezaev & Tregubova 2018).

Randall Collins visar att samtal oftast är regelstyrda i det att de vilar på en interaktionsordning som samtalsparterna delar. Därmed bör det vara möjligt för AI att tillägna sig en samtalskompetens. Harry Collins tillfogar till det att det mänskliga språket är dynamiskt, flexibelt och regelbrytande vilket innebär att samtalskompetens inte kan reduceras till kunskap om vilka regler och interaktionsordningar som vanligtvis gäller i en viss kontext eller situation. Med djup maskininlärning är det dock rimligt att anta AI kan utveckla en kontextuell kunskap inom vissa domäner. Precis som en individs tänkande utvecklas genom socialt samspel med andra personer och ting samt genom ett internaliserat samtal med sig själv, kan AI genom djup maskininlärning göra något liknande (jämför Goodfellow, Bengio & Courville 2016; Sejnowski 2018). En person kan indirekt – till exempel genom att läsa god skönlitteratur – tillägna sig andras erfarenheter och därmed öka sin förmåga till perspektivbyte, emotionell

15 På YouTube finns flera filmer där Apples digitala assistent Siri och Amazons motsvarighet Alexa får samtala med varandra. I vissa fall förstår assistenten frågan och ger ett relevant svar, i andra fall blir svaret ”I'm sorry, I'm afraid I don't have an answer to that”. Det intressanta här är att samtalet har en ytterst enkel turtagningsstruktur vilket få människor skulle acceptera i sin vardagliga konversation med andra människor.

medkänning och empati (Nussbaum 1998). AI kan kanske på ett liknande sätt tillägna sig en (viss) social intelligens och erhålla en icke-artikulerad (och kanske inte ens artikulerbar) kunskap. Själva input-sidan – tillgång till erfarenheter av vad som sker i sociala samspel – verkar därmed inte särskilja AI från människan. Däremot kan den tillägna sig och bearbeta denna erfarenhet på andra sätt och med andra resultat som följd. Det rör sig om generell intelligens men inom en specifik domän med dess specifika sociala och språkliga villkor. Däremot kanske AI inte kommer att utveckla social intelligens för andra domäner.[16]

Detta blir inte minst tydligt i Tegmarks diskussion om intelligens där han definierar den som förmågan att nå komplexa mål. Tegmark (2017:110–114) beskriver hur AI kunde lära sig nya spel och strategier genom att endast vara förprogrammerad att maximera sina poäng och sedan ständigt få sin aktuella poängställning uppdaterad. Denna typ av lärande – betingad djupinlärning (*deep reinforcement learning*) – har sedan överförts till en mängd andra områden än spel. Överfört till människans värld innebär det att AI kan utformas till att förstå det sociala livet som ett strategiskt och poänggivande spel och då själv utveckla nya förmågor i syfte att maximera sina poäng. Det handlar således om ett instrumentellt och målrationellt handlande.

För att bli en skicklig spelare i ett strategispel krävs att man i sina förberedelser och under spelets gång förstår hur ens motspelare tänker.[17] I denna bemärkelse kan troligtvis AI göra ett rollövertagande och kanske även utveckla en viss typ av självmedvetande där den distanserar sig från sin egen strategi och kritiskt analyserar sitt eget beslutsfattande och agerande (förstår sin verklighet utifrån ett mig). Den kan troligtvis övergå från lekfas till spelfas (Mead 1972:151) och se hur en vinnande strategi i ett spel är kontextberoende. Även om AI kan utveckla denna förmåga handlar det om ett begränsat självmedvetande (strategiskt handlande) i en välavgränsad kontext (spelliknande interaktion). Detta är i sig en imponerande utveckling av AI och därtill en utveckling som har långtgående sociala implikationer. Men det handlar om vad Habermas (1984:333) benämner strategiskt handlande: en social handling där man tar hänsyn till andra aktörers uppfattningar i syfte att uppnå ett mål. Men det är inte ett kommunikativt handlande som syftar till att nå en gemensam förståelse.

Som socialpsykologisk forskning visat är det en komplex process att utvecklas till en person, där såväl direkt som symbolisk interaktion är oerhört viktig för att utvecklas till en varelse som inte enbart reagerar och agerar utan även samagerar

16 Man kan här göra analogin till ett barn som tillbringat merparten av sin uppväxt framför en bildskärm och med begränsad tillgång till direkt sociala interaktion. Hon kommer troligtvis att ha tillägna sig en kompetens att vara en fullödig person i cyberspace: genom att ta del av andras samspel på nätet och själv delta i dem har hon tillägnat sig en social kompetens om hur man interagerar på nätet och kan avläsa subtila kommunikativa signaler som finns i olika typer av internetkonversation. Om hon däremot skulle möta en person ansikte mot ansikte skulle hon troligtvis ha stora problem att följa den interaktionsordning och samtalsrytm som gäller för denna typ av socialt samspel.

17 Jämför AI-forskares diskussion om ”omvänd betingad djupinlärning” (Tegmark 2017:346–358): genom att studera vilka beslut människor fattar kan AI troligtvis rekonstruera vilka värderingar och mål som styr dessa beslut.

(Cooley 1992; Mead 1972). Därtill väcks emotionssociologiska frågor, till exempel i vad mån empati (och därmed även självkännedom) krävs för att kunna tolka andras kommunikation och bli en fullgod partner i mänskliga samtal och mänskligt samagerande.[18] Frågan väcks om det finns social intelligens som inte kan översättas till algoritmer och omskapas till välavgränsade beräkningar (som NAND-grindar och artificiella neuronnätverk kan utföra minst lika skickligt som en mänsklig hjärna).[19] Harry Collins (2018) är ytterst tveksam till det och hävdar att även om det principiellt kan visa sig möjligt för AI att utveckla en social intelligens besitter den i dag endast en mycket snäv intelligens.

Kan AI agera?

En fråga av central betydelse är om AI kan ha agens, det vill säga handla. Socialpsykologen Johan Asplund (2002) anser att det inte är möjligt. Det som krävs är att den handlande varelsen ska veta att hon handlade, vilken handling hon utförde samt varför hon handlade på detta vis. Att utifrån en bearbetning av ett beräkningsbart underlag komma fram till en slutsats är därför inte agens.[20] Däremot använder sig människan ofta av hjälpmedel (instrument) i sitt agerande och agerar alltid på en scen (omgivning och omständigheter) som ger förutsättningar för hennes agerande. Men vare sig hjälpmedlen eller scenen äger agens, utan det är den handlande människan som tar dem i bruk när hon handlar.

Frågan väcks om djup maskininlärning innebär att AI har agens.[21] Djupinlärning handlar om att AI har förmåga till att självständigt lära sig saker och därmed även att nå en annan typ av intelligens än den som enbart är baserad på beräkningskraft. Brädspelet go anses vara ett av de mest intelligenskrävande spelen i världen eftersom förgreningsfaktorn gör att snäv beräkningskraft inte är tillräckligt för att bli en skicklig spelare (vilket var tillräckligt när schackdatorn

18 Det finns AI-forskare som hävdar att AI kan utveckla ett medvetande. För Tegmark, som är fysiker, är medvetande något som emergerar ur ett visst partikelarrangemang vilket innebär att även maskiner bör kunna utveckla ett medvetande (eftersom medvetandet är, med Tegmarks ord, substratoberoende). Vad som avses med medvetande är dock omstritt och det finns många definitioner. Tegmark är väl medveten om detta och väljer en icke-antropocentrisk definition där medvetande definieras som ”subjektiv upplevelse” och argumenterar för att medvetande är ”sättet som information känns när den bearbetas på vissa komplexa sätt” (Tegmark 2017:419).

19 En NAND-grind är en digital krets som baseras på boolesk algebra, i det här fallet ”Not And”. En mängd sammankopplade NAND-grindar kan tillämpa alla beräkningsbara funktioner. Artificiella neutronnät är människoskapade nätverk som liknar hjärnan (ett biologiskt neuronnät) och som baseras på självlärande algoritmer.

20 För Asplund visar inte Turingtestet att en maskin kan ha mänsklig intelligens eller inte. Vad det endast visar är ändamålsenligheten eller funktionsdugligheten hos en typ av teknisk artefakt (ett datorprogram). I sin bedömning av Searles kinesiska rum betonar Asplund att människan i det experimentet inte är en agent utan endast en mellanhand.

21 Asplund (2002:13) publicerade sin bok för snart 20 år sedan och han nämner explicit att han diskuterar ”gammal AI” men anger samtidigt: ”Jag har dock ingen anledning att tro att AI av senare datum [...] skulle behöva föranleda några justeringar i mina resonemang.”

Deep Blue år 1997 besegrade den regerande världsmästaren Garri Kasparov). Utöver strategiska kunskaper och beräkningskraft anser många bedömare att spelet go även kräver kreativitet och intuition. Skickliga spelare kan till exempel se vilka ställningar som är starka och svaga utan att artikulera varför de ser dem som det. När datorprogrammet AlphaGo år 2016 besegrade den 18-faldige världsmästaren Lee Sedol såg AI-forskare det som ett bevis för att AI kan utveckla en intelligens likt människans. I matchen gjorde AlphaGo ett drag som trotsade all tidigare erfarenhet men som 50 drag senare visade sig vara avgörande för att vinna matchen. Detta tolkar AI-forskare som att datorspelet gjorde ett drag baserat på intuition och kreativitet eftersom den inte i förväg kunde veta (räkna fram) varför den gjorde just detta drag (Tegmark 2017:114–118).

Poängen som AI-forskare här gör är inte att AI nu kan besegra en människa i extremt krävande strategiska spel. Det avgörande här är att Lee Sedol inte besegrades av en mänskligt förprogrammerad dator (likt Deep Blue) utan av ett datorprogram baserat i artificiella neuronnätverk som självständigt filtrerar information och därmed lärde sig saker utan mänsklig instruktion och fattade beslut som ingen människa hade instruerat eller väglett den att göra.

Det finns dock starka skäl till att inte se AlphaGos seger som en indikator på att AI kan nå mänsklig intelligens. Skälet till det är att det trots spelets svårighetsgrad ändå handlade om en extremt enkel kontext: strategiskt beslutsfattande i en strikt regelstyrd kontext innehållande endast en annan aktör (motspelaren) och där det finns ett väldefinierat tekniskt mål (att vinna) som båda aktörerna delar.

Men även om det inte rör sig om social intelligens i Harry Collins bemärkelse handlar det om beslut fattade utifrån förutsägelse, planering och handlingsalternativ. AlphaGo är därmed mer än ett hjälpmedel eller en scen eftersom den fattar beslut utan mänsklig instruktion. Däremot är det diskutabelt om det kan kopplas till intention i den bemärkelse som Asplund avser. Samtidigt väcks frågan om mening och intention enbart kan knytas till individer. En sociologisk ståndpunkt är att organisationer och makrosociala aktörer kan fatta beslut och utveckla mål och strategier och att dessa beslut inte kan härledas till dess enskilda medlemmar. En radikalisering av detta synsätt har gjorts av vetenskapssociologerna Michel Callon och Bruno Latour som hävdar att även icke-sociala entiteter kan ha agens (Callon 1984; Latour 2005). De använder sig av begreppet ”aktant” för att betona att även maskiner och artefakter – inte minst i form av nätverk – har agentskap eftersom de ”gör” någonting. Även om denna ståndpunkt är alltför långtgående, visar de hur tekniska system blir alltmer sammanflätade med människor och organisationer och därmed inte kan reduceras till att enbart utgöra möjligheter och begränsningar för mänskligt handlade.[22] Snarare än att ställa frågan om AI kan ha agentskap

22 Detta är utgångspunkten för den postfenomenologiska forskningen som utgår från att mycket av vårt görande numera sker genom tekniska system eftersom vi lever ett teknifierat samhälle (Idhe 1993, 2010; Selinger 2006). Detta påverkar inte bara vad vi gör utan vilka vi är eftersom vårt görande och varande är intimt sammanvävda.

bör frågan ställas på vilka olika sätt som nätverk av människor, organisationer och artefakter skapar agens.[23]

Behöver AI vara social?

Vi är redan i dag omgärdade av en typ av superintelligenta system som är autonoma i förhållande till den enskilde människan och starkt begränsar och beskär hennes makt och handlingsutrymme. Det sociala livet kännetecknas av att individer, grupper och organisationer är beroende av och ständigt interagerar med institutioner och system som är bortom deras kontroll och där de som befolkar systemet (professioner och positioner) är utbytbara.[24] Expertsystem är till exempel opersonliga system som inte nödvändigtvis tänker och lär som mänskliga individer gör. På övergripande samhällsnivå kan man se byråkrati och kapitalism som en form av superintelligens i bemärkelsen autonoma och ansiktslösa system med egna sociala logiker som inte enkelt kan härledas ur individuella val och beslut. En del av dessa system är oavsedda konsekvenser av mänskligt handlande medan andra system är intentionellt skapade men som med tiden utvecklat egna sociala logiker och mål. Givetvis finns skillnader mellan dagens system och den typ av superintelligenta AI-system som AI-forskare varnar för, inte minst att den senare är autonom (oberoende av mänskliga insatser för sin reproduktion). Samtidigt finns det många likheter, där frågor om systems styrbarhet, välvillighet och robusthet är centrala. Ett exempel är diskussionen om kapitalismens (kortsiktiga) robusthet och oförmåga att hantera många av dagens globala hållbarhetsutmaningar (se till exempel Scranton 2019).

Föreställningen om superintelligens baseras på att den har en generell intelligens på en betydligt högre nivå än människan. Denna inramning gör att samhället kommer i bakgrunden och den mänskliga intelligensen i förgrunden. En sociologiskt mer relevant inramning är att betona att samhället blir alltmer beroende av den typ av intelligens som AI besitter. Parallellt med att AI utvecklas till att kunna tillägna sig kontextuell kunskap inom vissa domäner utvecklas samhället till att bli alltmer baserat på och beroende av algoritmisk kunskap. Ju fler formaliserade och tydliga regler det utvecklas för en verksamhet och en profession desto enklare blir det för AI att ersätta människan. Digitaliseringen av samhället är inte bara bred (omfattar allt fler områden) utan även djup (påverkar dessa områdens karaktär) (Whitehead & Wesch 2012). Allt fler verksamheter är digitaliserade och har blivit en integrerad del av såväl offentliga och kommersiella verksamheter som våra vardagsliv (Just & Latzer 2017). Det handlar

23 I sin studie av mediekulturens förändring visar Mülhoff (2019) hur den mänskliga intelligensen blir ”fångad” i människa–maskinnätverk men att denna hybrida intelligens ofta framställs som en renodlad artificiell intelligens. Det som framställs som tekniska system är egentligen sociotekniska system där maskininlärning har möjliggjorts genom såväl tekniska landvinningar som sociokulturella förändringar (i hans studies fall en förändrad mediekultur). Se även Elliots (2019, kapitel 1) beskrivning av den intima sammanlänkningen av digitalisering, AI och sociala relationer.

24 Detta är en av Peter Kåhres (2009) huvudpoänger i sin Luhmann-inspirerade avhandling: att kunskap och intelligens inte enbart bör knytas till individer och mikroprocesser utan finns på systemnivå där de är oberoende av mänskliga aktörer.

inte längre om en enkel interaktion människa–maskin utan om ett samspel där digitalisering och algoritmstyrd datahantering inte enbart utgör en plattform för socialt liv utan blivit en integrerad del av detta liv (Elliott 2019; McCarthy 2017).

Som Latour (2005) understryker finns en ömsesidig påverkan mellan människa och ting: när man utvecklar en artefakt ändras även de som interagerar med den. Detta är något som Stilgoe (2018) betonar i sitt påpekande att ingen teknik är född smart utan den blir smart i samspel med människor och organisationer. I detta samspel finns en dubbelriktad socialisering: det är inte enbart AI som lär sig av och anpassar sig till människan utan människan lär sig av och anpassar sig till AI. Till exempel förväntas i dag många medborgare att vara en del av samhällets digitala struktur och bli digitala medborgare (Curran 2018; Hintz, Dencik & Wahl-Jorgensen 2019). Det innebär att i mötet mellan människa och AI sker en ömsesidig anpassning och lärande med följd att samtals- och interaktionsmönster ändras. Till exempel innebär en ökad mängd självkörande bilar inte enbart att dessa blir allt skickligare på att analysera och ta hänsyn till bilar med mänskliga förare utan även att de mänskliga förarna alltmer kommer att anpassa sig till självkörande bilars körsätt och trafikrytm.

Alltmer interaktion och kommunikation sker i och genom digitala nätverk. Det innebär att det finns en mängd sociala sammanhang, fyllda med samtal och samspel, baserade i kroppslös närvaro eller i form av virtuella kroppar (avatarer). Harry Collins (2018:57–73) betonar att kontextuell kunskap får man genom att vara inbäddad i ett socialt sammanhang men i takt med det sociala livets digitalisering kan denna sociala inbäddning i vissa sammanhang ske utan direkt kroppsligt deltagande.[25] En konvergens utvecklas mellan mänskliga och maskinella kommunikations- och interaktionsmönster, åtminstone på vissa områden, där AI blir mer människolik i sin interaktion samtidigt som människan blir alltmer AI-lik i sin. Detta kan i sin tur ha stora konsekvenser för människans socialitet och vår grundläggande förståelse av människa och socialt liv.[26] Att AI, vare sig i svag eller i stark version, inte omfattar en social intelligens innebär inte att AI därmed får en mindre påverkan på samhället och det sociala livet. I vissa fall kanske den får en större påverkan just genom att den inte besitter social intelligens och att människan anpassar sig till det.

25 Den mest långtgående relativiseringen av människans kroppslighet diskuteras av Bostrom (2017:379) som skriver att den kommande utvecklingen kan innebära att metoder utvecklas som kan ”hjälpa dem [människorna] att lägga bort sin dödliga skepnad helt och hållet genom att ladda upp sina medvetanden till ett digitalt substrat och låta sin befriade ande förkroppsligas i en utsökt välmående virtuell gestalt”. Denna radikala tanke förutsätter en substratoberoende syn på medvetande, vilket även Tegmark förespråkar: att medvetande har sin grund i ett visst partikelarrangemang och endast förutsätter en viss struktur på själva informationsbearbetningen (Tegmark 2017:377–406)

26 Se till exempel Turkle (2017) om hur digitaliseringen påverkar människans kompetens, inte minst hennes förmåga till empati, självreflektion och meningsfulla samtal.

Skapar AI-förväntningar fel fokus?

All ny teknik är inbäddad i sociotekniska föreställningar som innehåller visioner om framtiden och vilka möjligheter och faror som är förknippade med den (Jasanoff & Kim 2015). Berättelser om tekniska system och deras förväntade effekter är performativa eftersom de underlättar eller försvårar en viss utveckling (Boyd & Holton 2018; Wajcman 2017). Förväntningarna påverkar inte bara vilken teknik som utvecklas utan även sociala sammanhang som många gånger anpassar sig till dessa förväntningar. AI:s utveckling är extremt förväntansstyrt, där varje framsteg kan ses som en pusselbit till ett större genombrott. Forsythe (2002) hävdar att medan andra forskningsområden har imperativet ”publish or perish!” har AI-forskningen ”demo or die!”. Skälet till det är att många av dagens teknologiska uppvisningar och framgångar ses som embryon till och indikatorer på AI:s framtida förmåga.

Faran är därför inte enbart att AI:s potential underskattas utan även att den överskattas. Harry Collins (2018:203ff.) farhåga är inte singularitet (superintelligens) utan det han kallar ”underkastelse” (*surrender*); att samhället underkastar sig en AI-intelligens som tillmäts en större intelligens än vad den faktiskt besitter. Det innebär att främsta risken inte är AI:s utveckling – till exempel fäster Collins stor tilltro till att AI-styrda (”självkörande”) bilar på sikt kommer innebära en ökad trafiksäkerhet. Faran är i stället att samhället tillskriver AI en större intelligens än vad den faktiskt har, vilket får stora sociala implikationer. Resultatet blir att den intelligens som AI kan uppnå blir normerande och påverkar vad som ses som intelligens.[27] Det är troligtvis därför Harry Collins ägnar så stort utrymme åt att diskutera social intelligens: att motverka att naiva förväntningar knyts till AI. Faran är således inte i första hand AI:s utveckling utan främst samhällets användning av och anpassning till AI.

En annan fara är att en för snäv förståelse skapas av AI. Frågan om superintelligens är extrem i bemärkelse att den handlar om en ny typ av risk: ett tekniskt system (kroppslös intelligens) som i ett digitaliserat samhälle kan visa sig vara ostyrbar, okontrollerbar och ostoppbar (Thomas, Nafus & Sherman 2018).

En sociologisk utgångpunkt är att teknik aldrig är extern och autonom till samhället utan en del av ett sociotekniskt system befolkade av aktörer med vissa intressen och olika resurser och där maktförhållanden, normer och regelverk möjliggör eller försvårar viss teknikutveckling. (Bijker, Hughes & Pinch 2012; MacKenzie & Wajcman 1999). AI:s utveckling drivs framför allt av nio transnationella företag: USA-baserade Google,

27 Som exempel kan tas datautvinning (*data mining*) där AI genom mönsterigenkänning och beräkningsmetoder finner samband och trender i oerhört stora datamängder. Precis som sociologer och antropologer skapar kunskap om andra sammanhang än de man själva är en del av kan AI skapa kunskap om samhällsområden utan att vara en del av dem. Vetenskapssociologin betonar dock att all kunskap förutsätter en tolkningsgemenskap som kan avgöra tillförlitligheten i ett kunskapsanspråk (Collins & Evans 2007; Jasanoff 2005). Denna tolkningsförmåga erhålls genom gruppsocialisation (Lidskog & Sundqvist 2018). AI kan analysera en oerhörd mängd materia men frågan är på vilket sätt den bedömer tillförlitligheten i det material som den analyserar (vilket i vetenskapliga analyser sker genom att gruppspecifik expertis avgör vad som är tillförlitlig data och rimliga tolkningar av den).

Amazon, Apple, IBM, Microsoft och Facebook samt Kina-baserade Baidu, Alibaba och Tencent (Webb 2019). Samtidigt som AI diskuteras alltmer i samhället finner Webb (2019) att det finns få kritiska röster mot den ägarkoncentration som utmärker AI-företagen. Denna avsaknad av kritiskt perspektiv finns i Bostroms och Tegmarks diskussioner om AI:s sysselsättningseffekter och risken att teknikutvecklingen kan leda till ökad ojämlikheten. De betonar att det krävs en viss omfördelning av resurser så att en individs välstånd inte blir avhängigt av att hon har avlönad sysselsättning. Tegmark (2017:157) refererar till Brynjolfssons och McAffes (2011) metafor om ”ett digitalt Aten” där slavarna består av AI-robotar. Det finns dock ingen diskussion om vem som kommer utveckla och äga dessa AI-robotar och i vad mån ett gott liv enbart bör kopplas till materiellt välstånd eller även till makt att påverka sin livssituation.[28] Ur sociologiskt perspektiv är det viktigaste samtalet inte främst hur vi ska kunna styra en framtida superintelligens utan snarare vilka aktörer som styr dagens AI:s utveckling: vad sker i dag, varför sker det och vilka driver denna utveckling? Dessa djupt politiska frågor handlar om vilken samhällsutveckling vi ser som önskvärd och realiserbar (jämför Jasanoff 2003).

Avslutning: AI som system och föreställning

Allt tyder på att utvecklingen av AI kommer att fortgå och därmed bli ett allt viktigare område för sociologer att utforska. Samtidigt är det viktigt att undvika teknikdeterministiska och oreflekterat normativa perspektiv där en viss teknikutveckling ses som oundviklig eller tillmäts en ensidigt positiv eller negativ innebörd (Boyd & Holton 2018). Det finns en mängd exempel där forskning helt styrd av tekniska möjligheter och sociala förväntningar har lett till oavsedda och oönskade konsekvenser (Mlynař, Alavi, Verma m.fl. 2018). I stället är den sociologiska uppgiften att ha ett reflekterat, prövande och öppet perspektiv i studiet av AI och dess implikationer (Reed 2014). Detta innebär ett kritiskt-konstruktivt perspektiv som avtäcker de antaganden och förväntningar som AI-satsningar baseras på och som skapar utrymme för samtal om möjligheter och faror med AI. Denna diskussion är något som även flera verksamma AI-forskare välkomnar (Cavallo, Dario & Fortunati 2018; Etzioni & Etzioni 2016; Taddeo & Floridi 2018; Tegmark 2017).

I takt med samhällets digitalisering och AI:s utveckling kommer denna teknik att bli alltmer sammanflätad med samhället där tidigare gränser mellan människa, maskin och natur blir svårare att dra (Kimura 2017; Richardson 2015). Vetenskapssociologiska studier beskriver detta i termer av ”ontologisk översvämning”: laboratoriets landvinningar påverkar gradvis vår konstitutionella förståelse av mänskligt liv och vad det innebär att vara människa (Jasanoff 2005, 2018). AI:s utveckling och ak-

28 Här bör påpekas att Tegmark (2017:358) ställer frågan vilka ”vi” är när han ska diskutera vilka mål som AI bör omfatta och Bostrom (2017:203–204) nämner i förbigående ”det första agentproblemet”: när AI-utvecklare och beställare är olika personer (det andra agentproblemet handlar om hur en superintelligens kan kontrolleras).

törers förväntningar knutna till denna utveckling förändrar inte bara samhällets utformning och våra sociala praktiker och normer utan även vår förståelse av människa och samhälle. Av det skälet är det av stor vikt att sociologer inte bara studerar AI som socioteknisk system och infrastruktur utan även som föreställning. Precis som biologi och genetik har påverkat vår förståelse av vad liv är och kan bli, håller nu fysik och AI på att påverka vår förståelse av vad samhället och det sociala livet är och kan bli. Det innebär att frågor om AI:s utveckling är en såväl bred som djup samhällsfråga. Det är därför viktigt att sociologin deltar och påverkar det samtal som AI-forskare inbjuder till.

Referenser

Abrahamian, A.A. (2019) ”How the asteroid-mining bubble burst. A short history of the space industry’s failed (for now) gold rush”, *MIT Technology Review,* 26 juni 2019. https://www.technologyreview.com/s/613758/asteroid-mining-bubble-burst-history/

Asplund, J. (2002) *Genom huvudet. Problemlösningens socialpsykologi.* Göteborg: Bokförlaget Korpen.

Bainbridge, W.S., E.E. Brent, K.M. Carley, D.R. Heise, M.W. Macy, B. Markovsky & J. Skvoretz (1994) ”Artificial social intelligence”, *Annual Review of Sociology* 20:407–436. https://doi.org/10.1146/annurev.so.20.080194.002203

Bijker, W.E., T.P. Hughes & T.J. Pinch (red.) (2012) *The social construction of technological systems. New directions in the sociology and history of technology*. Cambridge, Massachusetts: MIT Press.

Bostrom, N. (2017[2014]) *Superintelligens. Vägar, faror, strategier.* Lidingö: Fri tanke.

Bostrom, N. & A. Sandberg (2017) ”The wisdom of nature. An evolutionary heuristic for human enhancement”, 189–219 i D. Ho (red.) *Philosophical issues in pharmaceutics. Development, dispensing, and use*. Dordrecht: Springer. https://doi.org/10.1007/978-94-024-0979-6_12

Boyd, R. & R. Holton (2018) ”Technology, innovation, employment and power. Does robotics and artificial intelligence really mean social transformation?”, *Journal of Sociology* 54 (3):331–345. https://doi.org/10.1177/1440783317726591

Braun, A., A. Zweck & D. Holtmannspötter (2016) ”The ambiguity of intelligent algorithms: Job killer or supporting assistant”, *European Journal of Futures Research* 4 (1):1–8. https://doi.org/10.1007/s40309-016-0091-3

Brent, E.E. (1988) ”Is there a role for artificial intelligence in sociological theorizing?”, *American Sociologist* 19 (2):158–166. https://doi.org/10.1007/BF02691809

Broadbent, S. (2016) *Intimacy at work. How digital media bring private life to the workplace*. Walnut Creek: Left Coast Press. https://doi.org/10.4324/9781315426136

Brougham, D. & J. Haar (2017) ”Employee assessment of their technological redundancy”, *Labour & Industry* 27 (3):213–231. https://doi.org/10.1080/10301763.2017.1369718

Brynjolfsson, E. & A. McAfee (2011) *Race against the machine. How the digital revo-*

lution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Lexington: Digital Frontier Press.
Carley, K.M. (1996) ”Artificial intelligence within sociology”, *Sociological Methods and Research* 25 (1):3–30. https://doi.org/10.1177/0049124196025001001
Callon, M. (1984) ”Some elements of a sociology of translation. Domestication of the scallops and the fishermen of St Brieuc Bay”. *The Sociological Review* 32 (1 suppl.):196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x
Cavallo, F., P. Dario & L. Fortunati (2018) ”Introduction to special section ’Bridging from user needs to deployed applications of social robots’”, *Information Society* 34 (3):127–129. https://doi.org/10.1080/01972243.2018.1444241
Collins, H. (1990) *Artificial experts. Social knowledge and intelligent machines.* Cambridge, Massachusetts: MIT Press.
Collins, H. (2018) *Artifictional intelligence. Against humanity’s surrender to computers*. Cambridge: Polity Press.
Collins H. & R. Evans (2007) *Rethinking expertise*. Chicago: The University of Chicago Press.
Collins, H. & T. Pinch (1998) *The Golem at large. What you should know about technology.* Cambridge: Cambridge University Press.
Collins, R. (2008[1992]) *Den sociologiska blicken. Att se bortom det uppenbara*. Lund: Studentlitteratur.
Cooley, C.H. (1992[1902]) *Human nature and the social order*. New Brunswick: Transaction Publishers.
Curran, D. (2018) ”Risk, innovation, and democracy in the digital economy”, *European Journal of Social Theory* 21 (2):207–226. https://doi.org/10.1177/1368431017710907
Dreyfus, H. (1972) *What computers can’t do. The limits of artificial intelligence.* New York: Harper & Row.
Dreyfus, H. (1992) *What computers still can’t do. A critique of artificial reason.* Cambridge, Massachusetts: MIT Press.
Elliott, A. (2019) *The culture of AI. Everyday life and the digital revolution.* London: Routledge. https://doi.org/10.4324/9781315387185
Etzioni, A. & O. Etzioni (2016) ”AI assisted ethics”, *Ethics and Information Technology* 18 (2):149–156. https://doi.org/10.1007/s10676-016-9400-6
Forsythe, D.E. (2002) *Studying those who study us. An anthropologist in the world of artificial intelligence*. Stanford: Stanford University Press.
Frey, C.B. & M.A. Osborne (2017) ”The future of employment. How susceptible are jobs to computerisation?”, *Technological Forecasting and Social Change* 114:254–280. http://dx.doi.org/10.1016/j.techfore.2016.08.019
Garfinkel, H. (1984[1967]) *Studies in ethnomethodology*. London: Polity Press.
Goffman, E. (1970[1967]) *När människor möts. Studiet av det direkta samspelet mellan människor.* Stockholm: Aldus/Bonnier.
Goodfellow, I., Y. Bengio & A. Courville (2016) *Deep learning.* Cambridge, Massachusetts: MIT Press.

Habermas, J. (1984) *The theory of communicative action. Vol. 1. Reason and the rationalization of society.* London: Heinemann.

Hintz, A., L. Dencik & K. Wahl-Jorgensen (2019) *Digital citizenship in a datafied society*. Cambridge: Polity Press.

Holton, R., & Boyd, R. (2019). "'Where are the people? What are they doing? Why are they doing it?' (Mindell) Situating artificial intelligence within a socio-technical framework." *Journal of Sociology*. Nätpublicering. https://doi.org/10.1177/1440783319873046

Ihde, D. (1993) *Postphenomenology. Essays in the postmodern context.* Evanston: Northwestern University Press.

Ihde, D. (2010) *Heidegger's technologies. Postphenomenological perspectives.* New York: Fordham University Press.

Jasanoff, S. (2003) "Technologies of humility. Citizen participation in governing science", *Minerva* 41 (3):223–244. https://doi.org/10.1023/A:1025557512320

Jasanoff, S. (2005) *Designs on nature. Science and democracy in Europe and the United States*. Princeton: Princeton University Press.

Jasanoff, S. (2018) *Can science make sense of life?* Cambridge: Polity Press.

Jasanoff, S. & S.-H. Kim. (red.) (2015) *Dreamscapes of modernity. Sociotechnical imaginaries and the fabrication of power*. Chicago: The University of Chicago Press.

Just, N. & M. Latzer (2017) "Governance by algorithms. Reality construction by algorithmic selection on the Internet", *Media, Culture & Society* 39 (2):238–258. https://doi.org/10.1177/0163443716643157

Kelleher, J. (2019) *Deep learning.* Cambridge, Massachusetts: MIT press.

Kimura, T. (2017) "Robotics and AI in the sociology of religion. A human in imago roboticae", *Social Compass* 64 (1):6–22. https://doi.org/10.1177/0037768616683326

Kåhre, P. (2009) *På AI-teknikens axlar. Om kunskapssociologin och stark artificiell intelligens*. Lund: Lunds universitet. https://portal.research.lu.se/ws/files/3898263/1415449.pdf

Latour, B. (2005) *Reassembling the social. An introduction to actor-network-theory*. Oxford: Oxford University Press.

Lee, A.J. & P.S. Cook (2020) "The myth of the 'data-driven' society. Exploring the interactions of data interfaces, circulations, and abstractions", *Sociology Compass* 14 (1): e12749. https://doi.org/10.1111/soc4.12749

Lidskog, R. & G. Sundqvist (2018) "Environmental expertise as group belonging. Environmental sociology meets Science and Technology Studies", *Nature and Culture* 13 (3):309–331. https://doi.org/10.3167/nc.2018.130301

Lupton, D. (2014) *Digital sociology*. Abingdon: Routledge. https://doi.org/10.4324/9781315776880

MacKenzie, D.A. & J. Wajcman (red.) (1999) *The social shaping of technology*. Buckingham: Open University Press.

Makridakis, S. (2017) "The forthcoming Artificial Intelligence (AI) revolution. Its impact on society and firms", *Futures* 90:46–60. https://doi.org/10.1016/j.futures.2017.03.006

McCarthy, M. (2017) ”The semantic web and its entanglements”, *Science, Technology and Society* 22 (1):21–37. https://doi.org/10.1177/0971721816682796
McClure, P.K. (2018) ”’You’re fired’, says the robot. The rise of automation in the workplace, technophobes, and fears of unemployment”, *Social Science Computer Review* 36 (2):139–156. https://doi.org/10.1177/0894439317698637
Mead, G.H. (1972[1934]) *Mind, self, and society. From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.
Mlynař, J., H.S. Alavi, H. Verma & L. Cantoni (2018) ”Towards a sociological conception of artificial intelligence”, 130–139 i M. Ikle, A. Franz, R. Rzepka & B. Goertzel (red.) *Artificial general intelligence 11th International Conference, AGI 2018, Prague, Czech Republic, August 22–25, 2018, Proceedings*. Cham: Springer. https://doi.org/10.1007/978-3-319-97676-1_13
Mülhoff, R. (2019) ”Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning”, *New Media & Society*. Nätpublicering. https://doi.org/10.1177/1461444819885334
Nasa (2020) ”Mission to a metal world: Psyche”. https://www.jpl.nasa.gov/missions/psyche/ (hämtningsdatum 15 april 2020).
Nussbaum, M. (1998) *Cultivating humanity. A classical defense of reform in liberal education*. Cambridge, Massachusetts: Harvard University Press.
Perdue, R. (2017) ”Superintelligence and natural resources. Morality and technology in a brave new world”, *Society and Natural Resources* 30 (8):1026–1031. https://doi.org/10.1080/08941920.2016.1264652
Polson, N. & J. Scott (2018) *AIQ. Hur artificiell intelligens fungerar.* Göteborg: Daidalos.
Rezaev A.V. & N.D. Tregubova (2018) ”Are sociologists ready for ’Artificial Sociality’? Current issues and future prospects for studying Artificial Intelligence in the social sciences”, *Monitoring of Public Opinion: Economic and Social Changes* (5):91–108. https://doi.org/10.14515/monitoring.2018.5.10
Reed, T.V. (2014) *Digitized lives. Culture, power, and social change in the internet era.* Abingdon: Routledge. https://doi.org/10.4324/9780203374672
Richardson, K. (2015) *An anthropology of robots and AI. Annihilation anxiety and machines.* New York: Routledge. https://doi.org/10.4324/9781315736426
Schwab, K. (2017) *The fourth industrial revolution*. Geneva: World Economic Forum.
Schwartz, R.D. (1989) ”Artificial intelligence as a sociological phenomenon”, *Canadian Journal of Sociology* 14 (2):179–202. https://doi.org/10.2307/3341290
Scranton, R. (2019) *Att lära sig dö i antropocen. Reflektioner över en civilisations slut*. Stockholm: Lil’Lit förlag.
Searle, J. (1980) ”Minds, brains and programs”, *Behavioral and Brain Sciences* 3 (3):417–457. https://doi.org/10.1017/S0140525X00005756
Sejnowski, T.J. (2018) *The deep learning revolution.* Cambridge, Massachusetts: MIT Press.
Selinger, E. (red.) (2006) *Postphenomenology. A critical companion to Ihde.* Albany: State University of New York Press.

Stilgoe, J. (2018) ”Machine learning, social learning and the governance of self-driving cars”, *Social Studies of Science* 48 (1):25–56. https://doi.org/10.1177/0306312717741687

Taddeo, M. & L. Floridi (2018) ”How AI can be a force for good”, *Science* 361 (6404):751–752. https://doi.org/10.1126/science.aat5991

Taylor, M.R., E.J Simon, J.L. Dickey, K. Hogan, J.B. Reece & N.A. Campbell (2017) *Campbell biology. Concepts & connections.* New York: Pearson Education Limited.

Tegmark, M. (2017) *Liv 3.0. Att vara människa i den artificiella intelligensens tid*. Stockholm: Volante.

Thomas, S., D. Nafus & J. Sherman (2018) ”Algorithms as fetish. Faith and possibility in algorithmic work”, *Big Data and Society* 5 (1):1–11. https://doi.org/10.1177%2F2053951717751552

Turkle, S. (2017) *Tillbaka till samtalet. Samtalets kraft i en digital tid.* Göteborg: Daidalos.

Vetandets värld (2018) ”De smarta maskinernas tid”, sänt i SVT2 27 augusti 2018. https://www.svt.se/nyheter/vetenskap/varldsberomda-fysikern-sverige-maste-satsa-pa-ai-innan-det-ar-for-sent

Wajcman, J. (2017) ”Automation: is it really different this time?”, *The British Journal of Sociology* 68 (1):119–127. https://doi.org/10.1111/1468-4446.12239

Webb, A. (2019) *The big nine. How the tech titans & their thinking machines could warp humanity.* New York: PublicAffairs.

Whitehead, N. & M. Wesch (red.) (2012) *Human no more. Digital subjectivities, unhuman subjects, and the end of anthropology*. Boulder: University Press of Colorado.

Woolgar, S. (1985) ”Why not a sociology of machines? The case of sociology and artificial intelligence”, *Sociology* 19 (4):557–572. https://doi.org/10.1177/0038038585019004005

Författarpresentation

Rolf Lidskog är professor i sociologi vid Institutionen för humaniora, utbildnings- och samhällsvetenskap, Örebro universitet. Han har sedan lång tid bedrivit forskning om expertisen roll för reglering av globala miljöproblem. För närvarande leder han projekten ”Hur skapas miljöexpertis? Institutionaliserad expertis, gränsorganisationer och globala miljöproblem” (finansierat av Vetenskapsrådet) och ”Att göra kunskap användbar. Interdisciplinära och transdisciplinära utmaningar för internationell miljöexpertis” (finansierat av Formas).

Kontaktuppgifter författare

Rolf Lidskog
Institutionen för humaniora, utbildnings- och samhällsvetenskap
Örebro universitet, 701 82 Örebro
rolf.lidskog@oru.se