



LUND UNIVERSITY

Sju nyanser av transparens

Om artificiell intelligens och ansvaret för digitala plattformars samhällspåverkan

Larsson, Stefan

Published in:
Plattformssamhället

2018

Document Version:
Förlagets slutgiltiga version

[Link to publication](#)

Citation for published version (APA):

Larsson, S. (2018). Sju nyanser av transparens: Om artificiell intelligens och ansvaret för digitala plattformars samhällspåverkan. I *Plattformssamhället: Den digitala utvecklingens politik, innovation och reglering* (s. 277-313). Fores.

Total number of authors:
1

Creative Commons License:
CC BY-NC-ND

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

STEFAN LARSSON

SJU NYANSER AV TRANSPARENS

OM ARTIFICIELL INTELLIGENS
OCH ANSVARET FÖR
DIGITALA PLATTFORMARS
SAMHÄLLSPÅVERKAN



PETERS *But you also know that artificial intelligence is not without its risk and that you have to be very transparent about how those algorithms are constructed. How do you see artificial intelligence, more specifically, dealing with the ecosystem by helping to get consumer insights, but also keeping consumer privacy safe?*

ZUCKERBERG *Senator, I think the — the core question you're asking about, A.I. transparency, is a really important one that people are just starting to very seriously study, and that's ramping up a lot. And I think this is going to be a very central question for how we think about A.I. systems over the next decade and beyond. Right now, a lot of our A.I. systems make decisions in ways that people don't really understand.*

PETERS *Right.*

ZUCKERBERG *And I don't think that in 10 or 20 years, in the future that we all want to build, we want to end up with systems that people don't understand how they're making decisions.¹*

¹ Utdrag ur Mark Zuckerbergs förhör vid USA:s senat, med frågor från senator Gary Peters, se exv. The Washington Post (10 april 2018) "Transcript of Mark Zuckerberg's Senate hearing".

INTRODUKTION: PLATTFORMAR, AI OCH ANSVAR

INOM SÄVÄL GLOBAL politik som samhällsvetenskaplig och rättsvetenskaplig forskning om digitala plattformar diskuteras alltmer hur man kan eller bör förstå de digitala plattformarnas ansvar för hur plattformarna används och deras samhälleliga konsekvenser.² Givet många av plattformarnas storskalighet, enorma användarbas och den stora mängd information som postas, modereras och sorteras – ofta i realtid – så har ett stort mått av mjukvarubaserad algoritmberoende automation blivit en del av organiseringen. Det betyder i sin tur att idéer om normer och behov av styrning och implementering av policy kommit att bli alltmer avhängiga varianter av artificiell intelligens och maskininlärning för sitt utförande. Ansvarsfrågorna för gränsdragningar och oförutsedda effekter av användningen av AI i autonoma system är ett alltmer omdebatterat fält i både nyhetsmedia och en växande mängd forskningslitteratur kring begrepp som *algorithmic accountability* och ansvarsfull AI. Tätt sammankopplat med frågor om ansvar är transparensfrågorna, dvs. hur man ska förklara ett visst utfall eller förlopp med autonoma, självlärande

2 Här kan bland annat Andersson Schwarz (2017), van Dijck et al. (2018) och Gillespie (2018) nämnas.

och automatiserade inslag, speciellt om man kombinerar det med proprietära företagsupplägg på komplexa datadrivna marknader.³

Givet att de storskaliga plattformarna behöver automatisera mycket av sin egen policyimplementering – i sina tjänsters funktionalitet och i innehållsmoderering – blir det därmed av stort intresse att förstå vad det medför i termer av direkta samhällsutmaningar. Hur ska vi förstå de etiska och sociala utmaningar som artificiell intelligens och maskininlärning medför i de digitala plattformarnas konsument- och marknadsinriktade användning? Hur påverkar det individer såväl som innovations förutsättningar, och finns det behov av reglering eller standardisering, och i så fall av vilken typ?⁴

Betydelsen av de utmaningar som följer av att digitala plattformar i många fall behöver ta en normativ ställning som ska implementeras genom automatiserade och självlärande verktyg ska inte bara ses i ljuset av de verksamheter som har en digital förhistoria, som med Google, Facebook och Amazon, utan även i ljuset av att verksamheter, som inte nödvändigtvis har en digital förhistoria, också anammar liknande datadrivna praktiker. Dvs, *plattformisering* kan beskrivas som en datafärd logik⁵ som kan hjälpa oss att bättre förstå de förändrade förutsättningar som verksamheter och företag – som i handel, media eller finans – ställs inför när de samlar in mer data som de analyserar med metoder som prediktiv analys och maskininlärning för att kunna individualisera tjänster eller effektivisera verksamheten.⁶ Kapitlets relevans ska alltså

3 Vilket det inledande citatet från Mark Zuckerbergs senatsförhör indikerar på. Frågan diskuteras i en växande mängd litteratur. Jag diskuterar det även i termer av en rättslig samhällsutmaning av relevans för fler användningsområden än för de digitala plattformarna, i Larsson (2019).

4 Även i en svensk politisk kontext har det på senare tid väckts frågor om regleringsutmaningar för både datadrivna marknader men också specifikt de algoritmstyrda utvecklingsområdena kring maskininlärning och artificiell intelligens. I maj 2018 publicerade exempelvis den svenska regeringen en *Nationell inriktning för artificiell intelligens*, som bl.a. innehåller en skrivning om att "Sverige behöver utveckla regler, standarder, normer och etiska principer i syfte att vägleda etisk och hållbar AI och användning av AI." (Regeringskansliet, 2018, s. 10).

5 Se Andersson Schwarz (2017).

6 Se även Andersson Schwarz & Larsson (2018).

inte enbart ställas i relation till de allra största och globala aktörerna.

Detta kapitel analyserar vad man kan kalla automatiserad policyimplementering på digitala plattformar och därmed kombinationen av normativitet och plattformisering. I kapitlet lägger jag extra fokus på betydelsen av de automatiserade verktyg där varianter av artificiell intelligens och maskininlärning används för att moderera och effektivisera olika informationsflöden. Jag tecknar tre kategorier där plattformar använder AI för att styra och reglera användare och andra aktörer; 1. en *infrastrukturell*, som framförallt rör de största när de skapar marknader för andra att agera på; 2. en *specifik*, gällande innehållsmoderering, och 3. en *generell*, i mötet mellan AI och normativitet, dvs. värdegrunder. Jag pekar då på några av de mest väsentliga utmaningarna med samhällsapplicerad AI, dvs. den användning som samspelar och reproducerar sociala strukturer och normer, och vad det innebär i termer av kompetensbehov inom AI avseende samhällelig bias, som kulturer, etik och sociala normer. Kopplat till ansvarsfrågorna diskuterar jag sju nyanser av transparens, i syfte att problematisera de ibland något förenklade kraven på radikal transparens som hörs i ansvarsutkrävande debatter och för att peka på hur marknadskomplexitet, och ägande samspelar med AI-metoderna till vad som kan kallas ”the black box”.

PLATTFORMAR

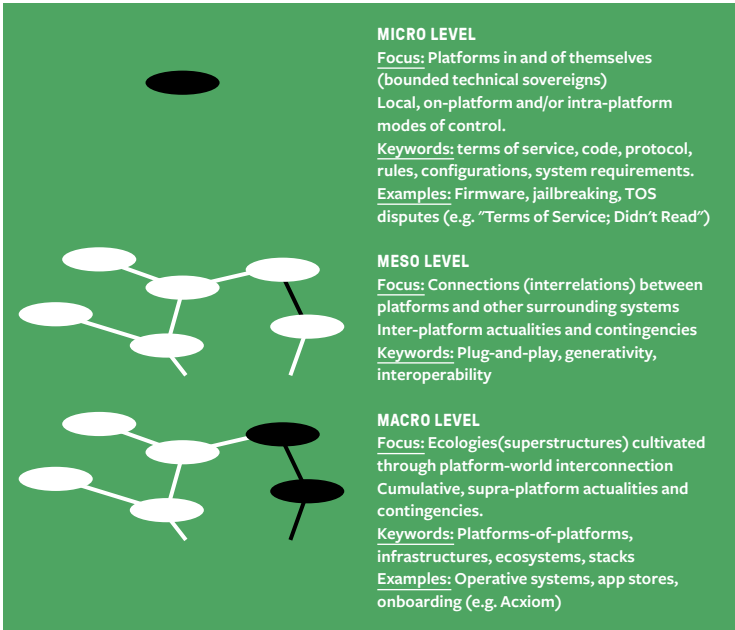
Först, vad är en digital plattform? I inledningskapitlet på antologin om *Plattformssamhället* pekar jag och Jonas Andersson Schwarz ut sju kvaliteter som är relevanta för definitionen av digitala plattformar. De är i hög grad:

1. mjukvarubaserade;
2. internetuppkopplade;
3. datafierande/datadrivna;

4. automatiskt/algorithmiskt sorterande i sin förmedling/distribution;
5. skalbara;
6. proprietära;
7. centraliserade.

Var och en av dessa kategorier är inte nödvändigtvis strikt uppfyllda, och det finns grader och komplexiteter inom vissa begrepp – som att delar av plattformars organisation kan vara väldigt öppna medan andra hårt proprietärt hållna, och att centralisering och decentralisering kan fördela sig på olika sätt. Även om alla dessa karaktäristikor är relevanta för att förstå digitala plattformars normativa utmaningar så är framförallt (3) *data driven* (4) *automatisering* i fokus här. Automatiseringen är i sin tur en förutsättning för (5) *skalbarheten*. De största plattformarna i termer av användare kan ha flera miljarder aktiva användare, och en moderering eller implementering av policy är inte möjlig utan en stor portion automatisering. I syfte att bedöma hur digitala plattformar har karaktären av informella rättssystem blir både (7) *centraliseringen* av hur policys tas fram viktig, precis som den svårighet som följer för granskning genom att essentiella delar av de interna strukturerna är dolda för externa granskare och där såväl kod som data är (6) *proprietär*, vilket gör ägandet och det proprietära upplägget relevantt att förstå betydelsen av i detta sammanhang. Både automationen, exempelvis i termer av maskininlärning, och det proprietära upplägget är, som jag utvecklar nedan, viktiga komponenter för att förstå vissa utmaningar med transparens.

Andersson Schwarz har i en vetenskaplig artikel från 2017 tecknat ett slags taxonomi över olika typer av plattformar, baserat på storlek och position i relation till varandra. En viktig del av att förstå plattformarnas informella styrmekanismer är att förstå denna ekologi, både i relationen mellan plattformar och individer och i relationen mellan storskaliga plattformar – infrastrukturella ”platform-of-platforms” – och mindre, men beroende, plattformsaktörer.



Plattformslogik. Från Andersson Schwarz (2017).

I den översikt som följer kommer jag att lyfta tre kategorier av rättsliga kvaliteter av relevans för att förstå utmaningarna med AI och normer hos plattformar, varav den första knyter an direkt till det ekologiska makronivåperspektiv Andersson Schwarz utgår från. På de nivåer där plattformar kopplas ihop med andra plattformar och bygger vidare på varandra (meso- och makronivåerna) finns en beroendeproblematik som sällan bara är informellt normativ utan rent hårdkodad, det vill säga den stipulerar konkreta villkor för olika aktörer att över huvud taget kunna göra affärer på den plattformiserade infrastrukturen.

PLATTFORMISERING

De två andra kategorierna jag utvecklar nedan har mer att göra med plattformarnas mikronivå; dels kan vi studera hur innehållsmoderering ständigt sker "ombord" på många typer av plattformar, dels kan vi se

hur den enorma skalan på flera av dessa plattformar kräver att plattformsföretagen tillgriper olika former av AI-lösningar, och att sådana lösningar i relation till människors språk och bilder och samhällsstrukturer överlag möter stora utmaningar i termer av inbyggda normer och samhällelig bias. Genom att teckna de digitala plattformarnas attribut kan man, som nämnt i inledningen av kapitlet, konstatera att också en rad traditionellt sett inte så digitala och automatiserat datadrivna branscher och aktörer anammar plattformisering. Exempel här är mediehus som Bonnier och Schibsted som, utöver att förmedla redaktionellt material, också bygger ett nätverk av datainsamlade sajter samtidigt som de i allt högre grad förmedlar individ- och gruppanpassad s.k. programmatisk och ”semantisk” reklam.⁷ Mediesajter är också den kategori som tycks ha mest spårningsmjukvara från tredjeparter på sina sajter.⁸ Även inom public service testas och diskuteras aspekter av skraddarsytt eller ”personaliserat” innehåll.⁹ En stor och ökande andel av handeln plattformiseras, både genom e-handel men även den fysiska handeln, genom datainsamling- och analys i klubb- eller lojalitetskort och utveckling av s.k. omnikanal.¹⁰ Praktiken blir också branschöverskridande när aktörer som ICA både har fysiska matvaru- och stormarknadsbutiker, digital försäljning, bank och försäkringsbolag. En omvänd historik kan man se hos Amazon, som primärt började som nätbokhandlare, men genom breddning av verksamheten till att erbjuda både molntjänster för företag, en handelsinfrastruktur som sammankopplar säljare med köpare, och uppköp av exempelvis matvarukedjan Whole Foods kommit att erbjuda en rad tjänster i enlighet med en i allra högsta grad plattformiserad logik. Liknande frågor kring datadrivenhet, automation och plattformologikens ekosystem kan diskuteras även inom skilda branscher som försäkring, bank och fastighetsförvaltning, vilket gör att följande analys om plattformars rättsliga kvaliteter och samhällets behov av transparens är relevant även för dessa verksamheter.

7 För mer om relationen mellan mediehus och nya digitala plattformsföretag, och den konfliktbild som finns, se Jonas Andersson Schwarz bidrag i denna volym. För mer om programmatisk reklam, se Larsson (2017b).

8 Englehardt & Narayanan (2016).

9 Se Jockum Hildéns bidrag i denna volym.

10 Larsson (2017b).

Användandet av användaravtal, nämnt som "terms-of-service" (Bild 1) är ett fält som studerats och problematiserats i en rad studier, framförallt i termer av att få läser, förstår eller tar ställning till dem.¹¹ Detta gör att de inte fungerar väl i en praktisk mening som ett instrument för transparens. Även om dataåtkomst, dvs. insamling av individers information, är en central förutsättning för mycket av det automatiserade policyarbete som plattformarna utför, så är fokus i detta kapitel på utmaningarna med automation, normer och styrning.

UTMANINGAR MED AUTOMATISERAD POLICYIMPLEMENTERING

Här tecknar jag tre kategorier där plattformars i hög grad automatiserade och maskininlärande aspekter används för att styra och reglera: 1. en *infrastrukturell*, av stor relevans för beroende andra aktörer; 2. en *specifik*, rörande moderering av innehåll, och 3. en *generell*, i mötet mellan AI och normer. Intresset ligger främst i att sortera i vilka typer av strukturella effekter som automatiserad policyimplementering har, för att kunna relatera till olika nyanser och behov av transparens från ett konsument- och samhällsperspektiv.

PLATFORMS-OF-PLATFORMS: INFRASTRUKTURELL POLICY

De största plattformarna har kommit att positionera sig som ett slags infrastruktur för en rad andra aktörer, som apputvecklare eller andra tjänsteutvecklare, vilket anger förutsättningarna för annan innovation, och i vissa fall andra branschens hela upplägg. Denna aspekt av governance eller typ av infrastrukturellt men aktivt styrande upplägg behöver inte i sig vara problematisk, men ställer stora krav på balansering

11 Se Per Lannerös bidrag i denna volym, men även Larsson (2018a; 2018b); Larsson & Ledendal (2017).

av egna och andras intressen, samt en praktik kring likabehandling och transparens som uppfattas som legitim hos de som påverkas. Exempel på detta är Apples App Store, som är enda vägen till att erbjuda appar för Appleprodukter. Varje app granskas av Apple i enlighet med en policy, vilket säkerställer kvalitet och teknisk funktionalitet, men som också har lett till kritik för en inkonsekvent hållning kring de delar av policyn som öppnar för tolkning i relation till kulturella värderingar eller i relationen mellan policy och dess implementering.¹² Sistnämnda tar Andersson Schwarz och jag upp i inledningen till denna volym i relation till svenska BankID, som egentligen bröt mot App Stores policy om att två appar inte fick vara beroende av varandra på det sättet, men framförallt ledde till ett ramaskri när policyn plötsligt implementerades skarpt och därmed stoppade BankID. Efter dialog släppte dock Apple fram BankID igen, till utvecklarnas stora glädje. Även informella rättssystem behöver utvecklas kontinuerligt, och när Apple 2010 skärpte upp implementeringen av sina policys träffades en rad appar med mer eller mindre lättklätt innehåll. En del kritik förde fram det inkonsekventa i att förbjuda baddräktsförsäljare som Simply Beach men tillåta Playboy och Victoria's Secret.¹³ Även om granskningen i sig inte är automatiserad så kan implementeringen av policyn vara det. Graden av automation i implementeringen tycks vara det som förändras över tid, vilket ger intressanta följd effekter som direkt påverkar de apputvecklare och företag världen över som utvecklar för användare av iPhone och iPad. Kontrollaspekten är central för Apples App Store, vilket tycks utmanas på den kinesiska marknaden. Apples App Store, som är den enda icke-kinesiske i Kina, tog, efter varning, bort omkring 700 appar för att deras uppdatering inte gått via Apples kontrollsystem.¹⁴ Orsaken till att den kinesiska marknaden utvecklar parallella uppdateringar till apparna är inte vederlagt, men det kan antas att det både handlar om geopolitik och statlig kontroll genom datainsamling genom apparna.¹⁵

12 Gillespie (2018).

13 Gillespie (2018, ss. 84-85).

14 The Telegraph (29 nov 2018).

15 Se Marina Svenssons bidrag om det kinesiska plattformssamhället i den här volymen.

Apples policyhantering över App Store handlar således både om att garantera funktionalitet, inklusive att undvika skadlig kod, men också att undvika hemlig tredjeparts inblandning. Upplägget, som möjliggör en appmarknad, leder också till ett beroende för apputvecklare för hur Apples policy ändras i sig eller i sin implementering. Vilket är ett utfall av den logik som följer upplägget med ”platforms-of-platforms”. Liknande beroendeställning har lyfts i förhållande till Facebooks ombalansering av den algoritm som styr det individuella nyhetsflödet, till att prioritera upp användarnas interaktioner på bekostnad av företags och nyhetsmediers poster, vilket bl.a. skedde i januari 2018.¹⁶ De strategiska val i viktandet av flödesalgoritmen som sker centralt hos Facebook har därmed direkta effekter för intäktsflödena och synligheten hos en rad kommersiella och andra verksamheter världen över.¹⁷

Googles sökmotor kan genom sin dominans på sökmarknaden också nämnas här, inte minst i ljuset av de beslut om höga viten för Google som EU-kommissionen tagit på konkurrensrättens område. I takt med att varumärkens synlighet blivit alltmer avhängigt hur ett företag eller en tjänst rankas av Googles sökmotorsalgoritm har också professionella kategorier och tjänster vuxit fram kring ”sökmotorsoptimering”, dvs. expertis kring att förhålla sig till sökmotorns interna relevansbedömningsupplägg i syfte att få en specifik sida eller tjänst att hamna högt upp i träfflistan. Det finns ett starkt ekonomiskt incitament eftersom användare i väldigt hög grad klickar sig vidare via de första träffarna. Här väcks också den konkurrensrättsliga frågan kring marknadsdominans när det gäller hur Google rankar och presenterar sökträffar för andra googletjänster, vilket bland annat aktualiserats i ett mångårigt fall prövat av EU-kommissionen. I juni 2017 utdömde kommissionen Google 2,42 miljarder euro – omkring 25 miljarder svenska kr – i böter

16 Business Insider (13 jan 2018) “Facebook is trying to prove it’s not a media company by dropping the guillotine on a bunch of media companies”, av Steve Kovach. <https://www.businessinsider.com/facebook-s-updated-news-feed-algorithm-nightmare-for-publishers-2018-1?r=US&lR=T&lR=T>

17 Se exv. TechCrunch (12 jan 2018) “Facebook feed change sacrifices time spent and news outlets for ‘well-being’” av Josh Constine. <https://techcrunch.com/2018/01/11/facebook-time-well-spent/>

för att bryta mot EU:s konkurrensrätt, vilket företaget har överklagat. Kommissionen menar att Google har missbrukat sin marknadsdominans som sökmotor genom att ge en annan Google-produkt, en prisjämförelsetjänst, en olaglig fördel gentemot andra tjänster av liknande slag.¹⁸

Samma problematik kopplad till sökmotorn finns i de krav Google ställer på de tillverkare av smartphones som använder Android, dvs. Googles operativsystem för smarta telefoner, och Googles appbutik (Play Store). I juli 2018 utdömde Kommissionen böter för Google på hissande 4,34 miljarder euro – omkring 48 miljarder svenska kr – för olagliga rutiner avseende Android-enheter som kommissionen menade otillbörligt stärker och bibehåller dominansen hos Googles sökmotor. Kommissionen konstaterade att Google har krävt att tillverkare förinstallerar appen Google Search och webbläsaren (Chrome), som ett villkor för licensiering av Googles appbutik (Play Store) och att Google hade gjort betalningar till vissa stora tillverkare och mobiloperatörer under förutsättning att de bara förinstallerade Google Search-appen på sina enheter.¹⁹ Googles vd, Sundar Pichai, konstaterade samma dag beslutet kom att de kommer att överklaga, och menade att kommissionen missar hur många valmöjligheter Android erbjuder telefontillverkare såväl som apputvecklare och konsumenter.²⁰

Samtidigt kommer uppgifter från flera håll om att Google utvecklar en kinaanpassad variant av sökmotor kallad ”Dragonfly”, som därmed skulle behöva vara censurerad för att tillåtas av den kinesiska staten.²¹ Vilket inte bekräftas av Google, och överhuvudtaget är svårt att bekräfta, och är en kontroversiell fråga. Googles sökmotor har dock tidigare funnits på den kinesiska marknaden, men krav på statligt kontrollerad censurering, inklusive en hackerattack riktad mot Google, fick

18 EU-kommissionen (27 juni 2017) Press release. http://europa.eu/rapid/press-release_IP-17-1784_en.htm

19 EU-kommissionen (18 juni 2018) Press release. http://europa.eu/rapid/press-release_IP-18-4581_en.htm

20 Pichai, S. (18 juli 2018) “Android has created more choice, not less”, Google in Europe. <https://www.blog.google/around-the-globe/google-europe/android-has-created-more-choice-not-less/>

21 The Intercept (29 nov 2018).

företaget att flytta sökverksamheten till det mindre hårt kontrollerade Hong Kong, vilket i förlängningen ledde till att vare sig Gmail, Chrome eller sökmotorn har funnits tillgängliga för större delen av den kinesiska marknaden sedan 2014.

Frågan om den kinesiska marknaden är dock intressant ur ett automatiserat policyperspektiv på flera sätt. Dels ser de globala men västliga plattformsgiganterna världens största interna marknad som något attraktivt, och dels har de kinesiska företag som på kort tid vuxit fram där kommit att börja konkurrera utanför Kina. Man skulle därmed kunna se Kina som en statligt kontrollerad, auktoritär och gigantisk testbädd för utveckling av allt från handelsplattformar till varianter av artificiell intelligens baserad på stora mängder tillgänglig data (bl.a. från publika kameror och sociala medier). I förlängningen ger det kinesiska företag ett slags konkurrensfördel gentemot västliga företag, som behöver följa andra spelregler, vilket troligen ses som ett reellt marknadshot av sistnämnda. Kanske behöver betydelsen av de västligas olika initiativ riktade mot den kinesiska marknaden – trots att den måste innebära en anpassning till den statligt auktoritära kontrollen – vägas mot just det hotet.²²

INNEHÅLLSMODERERING

Alla sociala medieplattformar använder varianter av mänsklig granskning, användarflaggning och automatiserat upptäckande.²³ Även om sociala medieplattformar bara är en delmängd av de aktörer som vi åsyftar när vi diskutera digitala plattformar överlag så är innehållsmodereringen ett särdeles viktigt organisatoriskt upplägg att studera för att förstå de utmaningar digitala plattformar står inför när det gäller att ta normativa beslut. Medieforskaren Tarleton Gillespie argumenterar i boken *Custodians of the Internet* (2018) att innehållsmoderering i själva verket är kärnan, en betydande del av definitionen, av sociala medieplattformar.

22 Se exempelvis diskussionen i *The New York Times* (2018-11-18) "China Rules. How China became a superpower", av Raymond Zhong. <https://www.nytimes.com/interactive/2018/11/18/world/asia/china-internet.html>

23 Se exv. Gillespie (2017).

Givet alla de hemskheter och illdåd som postas på sociala medieplattformar – tänk barnporr, halshuggningar och uppvigling till folkmord – så är det inte längre något kontroversiellt konstaterande att plattformarna behöver modereras. En fråga är dock vilket ansvar som plattformarna bör ha för upptäckten och lagföringen av de som utför illdåd som plattformarna därmed ser, och den svårighet som medföljer normativt beslutsfattande. Man kan konstatera att svårighetsgraden i att moderera är oerhört hög för gränsfall, kontextuella motsättningar, politiska meningsmotståndare som använder flaggningsfunktioner för att obstruera motståndare eller andra upptänkliga sätt att missbruka plattformarnas upplägg, utöver oavsiktliga effekter av en automatiserad implementering av policy. Man hamnar snabbt i utmaningar där man inte bara ska definiera vad som är oacceptabelt, utan även navigera i en mängd kulturer, kolliderande rättsordningar och laddade frågor, vilket Gillespie beskriver som:

... balancing offense and importance; reconciling competing value systems; mediating when people harm one another, intentionally or otherwise; honoring contours of political discourse and cultural taste; grappling with inequities of gender, sexuality, race, and class; extending ethical obligations across national, cultural, and linguistic boundaries; and doing all that around the hottest hot-button issues of the day.²⁴

Utmaningarna för de digitala plattformarnas arbete med automatisering och policyimplementering ligger därmed i en blandning av kontext, omtvistade fakta, och kombinationen av hastighet och skala. Speciellt realtidsmoderering ställer rimligen oerhörda krav på automatiserat upptäckande av innehåll som tydligt bryter mot användarpolicy. Dessa svårigheter gör också att frågan kring hur man inom utvecklingen av AI och användandet av maskininlärning ska hantera uppgifter av normativ och kontextuell karaktär. I en databeroende maskininlärning finns det ett beroende av gyllene standards för lärande av den modell man vill ska

24 Gillespie (2018, s. 10).

råda, dvs. ett slags konservatism. Frågan är vad denna typ av konservatism medför för utmaningar, inklusive oavsiktliga konsekvenser när den appliceras av storskaliga digitala plattformar.

En vidare diskurs med AI-relevans för utmaningar med ansvarsfördelning rör digitala plattformars betydelse, vilket delvis är en debatt om hur intermediärer ska bedömas när det gäller ansvar för det material eller beteende som sprids eller genereras i relation till plattformarna. Intermediärsansvarsfrågor är inget nytt i digitala sammanhang²⁵ men aktuella exempel kan man finna hos de storskaliga digitala plattformarna, som i debatterna kring vilket ansvar Facebook och YouTube (Google) bör ha för den information som delas inom deras respektive plattformar eller gällande Googles relevansbedömning i den indexerande sökmotorn²⁶. Eftersom plattformarna är storskaliga – Facebook har över två miljarder aktiva användare och Google har enligt uppgift inte mindre än sju tjänster som har mer än en miljard användare – behöver de också i hög grad automatisera informationshanteringen, där båda aktörerna är stora investerare i och utvecklare av artificiell intelligens för en rad funktionaliteter, bland annat för ansiktigenkänning, språkanalys och röstigenkänning²⁷.

Det finns en kritik mot YouTube som går ut på att den reklamäklade affärsmodellen bidrar till både sensationalism²⁸ och radikalisering. Den kritiske tekniksociologen Zeynep Tufekci konstaterar frankt att YouTube därmed kan vara ett av de mest kraftfulla radikaliserande instrumenten i det 21:a århundradet.²⁹ Liknande kritik har riktats mot Facebook i relation till dold riktad reklam av den amerikanske kulturhistorikern och medieforskaren Siva Vaidhyanathan, som menar att oavsett vem som låg bakom den stora mängd riktad reklam med politiska

25 När individerna bakom fildelningssajten The Pirate Bay åtalades 2009 för medhjälp till upphovsrättsbrott uppkom en liknande konceptuell utmaning kring hur domstolen skulle se på ansvar i förhållande till denna ”plattform”; se Larsson (2017a).

26 Jfr. Gillespie, 2018)

27 Dolata (2017)

28 Gillespie (16 jan 2018).

29 Tufekci (10 mars, 2018).

budskap inför det amerikanska presidentvalet 2016 så har Facebook likväl ett ansvar för vad upplägget möjliggör. Den del av valpåverkan som upplägget medverkade till var, enligt Vaidhyanathan, alltså inte ett missbruk av plattformen utan att *bruk av plattformen*, helt i enlighet med både design och affärsmodell.³⁰

En variant av intermediärsansvarsfrågan rör kontroll över användarinformation, vilket aktualiserades i den s.k. Cambridge Analytica-skandalen där information från mellan 50 till 87 miljoner Facebookanvändare visats ha använts för att påverka demokratiska val i en rad länder.³¹ När Facebooks VD, Mark Zuckerberg, frågades ut av den amerikanska kongressen med anledning av skandalen fick han frågor om plattformens ansvar för det material som sprids. Zuckerberg förde återkommande fram AI som ett verktyg för att bekämpa oönskat innehåll som hatiska uttalanden, falska nyheter, hämndporr och annat. Hans uttalanden har kritiserats för att vara uttryck för en förenklad AI *solutionism* och att de automatiserade optimeringsverktygen som den storskaliga plattformen bygger på i sig själva bidragit till spridningen av falska nyheter och kontroversiellt innehåll.³² En ansvarserkännande plattformdesign – kan konstateras – ställs inför en rad normativa ställningstaganden kring vilken typ av bilder, texter och länkar som ska räknas som stötande, olagliga eller falska. Inte sällan dras gränserna på olika sätt i olika kulturer och jurisdiktioner. Även en del kunskapsområden, exv. om historiska händelser eller geografiska platsers tillhörighet, kan vara kontroversiella och bestridas av endera grupper.

30 Vaidhyanathan (8 sept 2017).

31 En nyhet som fick mycket spridning när journalisten Carole Cadwalladr publicerade en artikel med en visselblåsare i The Guardian (18 mars 2018). <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump>

32 BuzzFeed (11 april 2018) "Why Facebook Will Never Fully Solve Its Problems With AI" av Davey Alba. <https://www.buzzfeednews.com/article/daveyalba/mark-zuckerberg-artificial-intelligence-facebook-content-pro> ; se även debattartikel i The New York Times (20 okt 2018) skriven av en professor i psykologi och neurovetenskap och en professor i datavetenskap, "No, A.I. Won't Solve the Fake News Problem. At least not for decades to come. Sorry, Mark Zuckerberg." Av Gary Marcus och Ernest Davis. <https://www.nytimes.com/2018/10/20/opinion/sunday/ai-fake-news-disinformation-campaigns.html>

Gillespie menar att innehållsmodereringen på sociala medieplattformar är en lika svår som behövlig uppgift. Svårigheterna med automatiserat och maskinlärande upptäckande av stötande poster har enligt Gillespie att göra med ”kulturens flyktighet, språkets komplexitet och anpassningsförmågan hos brottslingar som vill undvika upptäckt”.³³ Det betyder att det inte bara är avsiktligt missbruk som är utmaningen utan även den normativa karaktären av att implementera policy över en myriad av kulturella grupperingar och ett hundratal jurisdiktioner. Olika grupper kommer att ha olika kulturella standards och sociala normer, och en del av dessa är rentav direkt motstridiga andra.³⁴ Ett exempel handlar om hur man från ett designens policyperspektiv bör dra gränsen för nakenhet, som är ett fält där acceptansen och de sociala normerna ser olika ut i olika delar av världen, även för olika grupper inom samma samhälle. Facebooks hantering av amningsbilder har lett till protestmarscher, vilket i sin tur har lett till ett slags intern policyutveckling på detaljnivå kring hur gamla bebisarna får se ut att vara, hur mycket av bröstet som får synas, m.m. (såvitt vi vet – för det är ingen enkel sak att studera de interna policybesluten).³⁵ Den omtalade bilden på en naken napalmskadad vietnamesisk flicka – tagen av 1972-års Pulitzerprisvinnare Nick Ut – skapade förvisso kontroverser redan när den publicerades i anslutning till Vietnamkriget, men skapade återigen rubriker när Facebook tog bort poster med bilden på plattformen. Det var troligen en kombination av underårig nakenhet och stötande bild på lidande som träffade Facebooks policy och som därmed ledde till borttagande av poster, även om bilden ingick i den norska journalisten Tom Egelands artikel i Aftenposten, som reflekterade över fotografier som förändrat historien. Utmaningen här ligger i *kontextualitet*, dvs. att just denna bild har en historisk ställning som lett till tidigare publiceringsbeslut, även om de också utmanat rådande pressetik, där behovet av förståelse av hemsgheten i pågående krig har bedömts övertrumfa värdet av att inte publicera stötande bilder.³⁶

33 Gillespie (2018, s. 107): “The fluidity of culture, complexity of language, and adaptability of violators looking to avoid detection”.

34 Se även utmaningarna tecknade i Larsson (2018c).

35 Lysna exv. på Radiolabs “Post no evil” (17 aug, 2018). <https://www.wnycstudios.org/story/post-no-evil>

36 Jfr. Gillespie (2018, kap. 1).

AI, ANSVAR OCH NORMER

Centralt för detta kapitel är kombinationen av automatiserade processer och normativitet, vilket exempelvis är en del av den innehållsmoderering beskriven ovan. Givet skalbarhetens behov av automation så har också de digitala plattformarnas tillväxt kommit att länkas ihop med användandet av varianter av AI och maskininläring som en metod för policyimplementering, till exempel för att upptäcka missbruk eller användarbeteende som strider mot plattformarnas policy. Här fokuseras därför kombinationen av normativitet och AI i vid mening genom att ta upp några problematiska exempel ur den forskningslitteratur som växer fram i fältet för *Fairness, Accountability and Transparency (FAT)*.

I inledningskapitlet till antologin om Plattformssamhället argumenteras för behovet av att förstå plattformar som en samhällsfråga, vad utvecklingen åt att i ökande grad organisera samhällskritiska funktioner utefter en digital plattformslögik medför.³⁷ Där tecknas en lista över vilka attribut som digitala plattformar står för och där automation är centralt, som en följd av storskalighet och snabbhet i moderering. Det betyder i sin tur att artificiell intelligens, framförallt maskininläring, blir en metod- och teknologiutvecklingsfront som medföljer plattformarnas tillväxt. Intresset för maskininläring och varianter av AI är en del av en trend hos några av de största plattformsbolagen, inklusive Amazon, Apple, Google (som exempelvis förvärvade DeepMind 2014), Microsoft och Twitter, som köper AI-startups och investerar i egna studier.³⁸ Det betyder att en rad samhälleligt centrala funktioner – som påverkar arbetslivet, familjers ekonomiska förutsättningar, hur nyheter och kunskap distribueras och individers sjukvård – medieras i allt högre grad genom varianter av artificiell intelligens och enligt en digital plattformslögik. Detta ger att några av de rättsliga utmaningar som gäller

37 Se även Andersson Schwarz & Larsson (2018).

38 American Scientific (20 juni 2016) "When Will Computers Have Common Sense? Ask Facebook" av Larry Greenemeier. <https://www.scientificamerican.com/article/when-will-computers-have-common-sense-ask-facebook/>

digitala plattformar och plattformisering sammanfaller med behovet av att teckna och förstå de rättsliga utmaningar som gäller för en samhällsapplicerad artificiell intelligens.³⁹ Facebook använder exempelvis AI för att administrera sin plattform, som att identifiera sina användares individuella intressen och skraddarsy innehållet genom att automatiskt analysera deras nyhetsflöden och värdera interaktion och klick, identifiera personer i foton och rikta marknadsföring mot användarna i enlighet med deras intressen och preferenser, och ge förslag på vilka de borde bli vänner med på plattformen. Det innebär dock att Facebook också har en gigantisk databas med ansiktsgenkänningsrelaterad data, som kräver stor tillit till att den är säkert hanterad och inte missbrukas. Liknande gäller rimligen för Apples iCloud, Microsofts LinkedIn och Googles Photo.

En stor utmaning är också att undvika att falska konton startas som kan användas för att sprida propaganda eller nyheter som har som syfte att manipulera debatter, eller för att övervaka eller trakassera andra konton. Enligt Facebooks senaste s.k. *Transparency report* sägs att Facebook tog ned 754 miljoner falska konton i tredje kvartalet 2018. Och för de allra flesta av dessa, dvs. 99,6%, användes AI för nedtagandet innan en användare rapporterat dem och inom några minuter från det att de skapats.⁴⁰ Det är ett tydligt uttryck för en automatiserad policyimplementering som kräver avancerade varianter av artificiell intelligens som kan lära sig att känna igen och sortera ut sant från falskt, och som har en tydlig nytta för både användare, Facebook och reklamköpare.

Dock uppstår med automationen ibland kulturellt känsliga utfall, som kan vara oavsiktliga men ändå stötande. Ett exempel är när annonskategorier per automatik hämtades från användares intresseprofiler, vilket ledde till att antisemitiska kategorier erbjöds reklamintressenter som kategorier att rikta reklam mot, eftersom det helt enkelt fanns en grupp som tydligt i sina profiler angav att de hatade judar. Skandalen var ett

39 Se Larsson (2018b).

40 Facebook Transparency Report: <https://transparency.facebook.com>

faktum, och i den delen av plattformsadministreringen har nu kontrollmekanismer införts.⁴¹

Bildanalysen är avancerad inom en rad digitala plattformar, men även språkanalysen, och en trend har med ljud, dvs. tal, att göra. Apples Siri, Google Assistant, Amazons Alexa, och Microsofts Cortana är alla exempel på virtuella assistenter, där flera varianter utvecklats för att vara en interaktiv assistent i hemmiljö, som kan svara på frågor eller rentav i vissa fall kan beställa hem produkter. Det betyder också att hemmets sfär är en del av en strategisk satsning hos flera av företagen, där dataifiering av hur familjekonstellationer och hemmets konfigurering ser ut tycks vara en del, och konsumentnyttan ofta beskrivs i termer av att ha ett tal-interface för vardagliga uppgifter i hemmet.

Inom den kritiska och samhällsvetenskapligt förankrade AI-forskningen finns det framväxande insikter som har att göra med hur AI, applicerad av digitala plattformar, har samhällsliga baksidor och medför etiska och normativa utmaningar. Detta uttrycks ofta i termer av behovet av att säkerställa *fairness* eller opartiskhet i samhällsappliceringen av data- och algoritmdrivna system, dvs. att de inte bidrar till diskriminering eller samhällslig bias eller har andra framförallt oavsiktliga och oetiska konsekvenser.⁴² Detta knyts ofta till diskussioner om ansvarsfördelning eller ansvarserkännande (*accountability*) – till exempel riktat mot digitala plattformar – och behovet av transparens, i syfte att bättre kunna upptäcka och motverka negativa konsekvenser. I denna typ av forskningslitteratur poängteras att algoritmiska system används i ett antal sammanhang som med hjälp av stora datamängder ("big data") filtrerar, sorterar, betygsätter, rekommenderar, "personifierar" och på andra sätt formar

41 ProPublica (14 september 2017) "Facebook Enabled Advertisers to Reach 'Jew Haters'" av Julia Angwin, Madeleine Varner and Ariana Tobin. <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>

42 När det gäller rättsliga, etiska och sociala frågor kring automation genom artificiell intelligens som samspelar med individens data så kan institutet AI Now nämnas, med Kate Crawford i spetsen, och institutet Data & Society med danah boyd som föreståndare. IEEE har genom en extensiv samling mångvetenskapliga forskare också tagit fram goda rekommendationer att fortsätta bearbeta i *Ethically Aligned Design* (v.2, 2018). För en svensk framställning, se Larsson (2019; 2018c).

mänskliga erfarenheter och förhållanden. Även om dessa system ger många fördelar, innehåller de också inneboende risker, såsom kodifiering och förstärkande av samhällelig bias, reducerad ansvarsskyldighet, och ökad informationsasymmetri mellan dataproducenter (kunder) och datainnehavare.

Exempelvis, en vetenskaplig utvärdering av tre kommersiella könsbestämmande bildigenkänningsystem visade att kvinnor med mörkare hy är den grupp med störst grad av felklassificering.⁴³ Det betyder bland annat att dessa tjänster, och applikationer som bygger på dessa, fungerar sämre för dessa gruppers utseenden. Delar av orsaken tycks ha att göra med att de underliggande bilddatabaserna är överrepresenterade av vita män, vilket leder till att felmarginalen är påtagligt mycket mindre för ljushyade män. På applikationsnivå kan den här typen av bias leda till att kamerafunktioner som ska tipsa fotografen när den fotograferade blinkar tolkar det som att asiater alltid blinkar⁴⁴ eller att en bild-databas automatiskt taggar svarta människor som ”gorillor”.⁴⁵

I boken *Algorithms of Oppression* argumenterar kultursociologen och informatikforskaren Safiya Noble för att sökmotorer har en roll i att inte bara reproducera utan även förstärka de rasistiska strukturer som finns i samhället – hon inriktar sig mot afroamerikaner i USA.⁴⁶ En utmaning här tycks vara att relationen mellan samhällets inneboende strukturer och historik av både ojämställdhet och ojämlikhet – kommer att representeras i den data som lärande algoritmer tränas på. Om samhällelig bias finns i datan replikeras det i utfallet. När det gäller bilddatabaser konstaterar Zou & Schiebinger i en kommentar i *Nature*

43 Buolamwini & Gebru (2018).

44 Vilket kommenteras av forskare i *Nature*, se Zou, J. & Schiebinger, L. (18 juli 2018).

45 Vilket fick Google, som utvecklat tjänsten, att helt ta bort tjänstens möjlighet att automattagga någonting som gorilla. Detta konstaterar Wired efter att ha testat mer än 40 000 djurbilder i tjänsten. Andra djur kan fortfarande automattaggas. Detta kan tolkas som att bildigenkänning fortfarande har en del stora utmaningar. Se exv. The Guardian (12 januari 2018) “Google’s solution to accidental algorithmic racism: ban gorillas”, av Alex Hern. <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people>

46 Noble (2018).

att neurala nätverk för sin bildklassificering ofta tränas på en databas som heter ImageNet. Även om databasen innehåller mer än 14 miljoner klassificerade bilder så kommer de allra flesta från ett fåtal länder.⁴⁷ I ett annat exempel upptäckte datavetenskapliga forskare vid University of Virginia att populära bilddatabaser innehöll genderbias, där kvinnor kopplades till köket och män till jakt, vilket ledde till en maskininlärande applikation som inte bara reproducerade utan även förstärkte samma bias.⁴⁸ Detta leder till en fråga som inte bara handlar om att fylla på med mer representativ data, utan vilket ansvar den som designar applikationer som använder sig av samhällsligt skev data har för att uppmärksamma och eventuellt motverka denna skevhet.

” Ska datan vara representativ för världen som den är, eller för en värld som många aspirerar mot? Utmaningen ligger i relationen till samhällets strukturer. Representativ data kan ju de facto beskriva ett samhälle som innehåller diskriminerande praktiker, ojämlik lönesättning, sinsemellan oförenliga normativa uppfattningar om familjers konstitution, hur relationer bör få se ut, och etnisk, religiös eller könsbaserad diskriminering.

Utmaningen ligger i relationen till samhällets strukturer, där den representativa datan de facto kan beskriva ett samhälle som innehåller diskriminerande praktiker, ojämlik lönesättning, sinsemellan oförenliga normativa uppfattningar om familjers konstitution, hur relationer bör få se ut, med etnisk, religiös eller könsbaserad rasism och diskriminering. Frågan om legitimitet och fairness blir därmed desto svårare. Zou

47 Zou, J. & Schiebinger, L. (18 juli 2018).

48 se Zhao et al. (2017) för studie. Studien rapporterades om i tidningen Wired 21 augusti 2017: <https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/amp>

& Schiebinger konstaterar därmed kärnfullt: "ska datan vara representativ för världen som den är, eller för en värld som många aspirerar mot?".⁴⁹ Oavsett val här, från ett designperspektiv, så blir frågan normativ. Och, i linje med den utmaning kring kontextualitet jag lyfter ovan, frågar de sig vilka föreställningar om "fairness" eller opartiskhet som ska prioriteras. I många fall kommer rimligen ingen konsensus att gå att nå. Oavsett vilket, frågan om ansvarserkännande för digitala plattformars användning av AI i relation till människor och kulturella strukturer vaknar här någonstans, i takt med att en neutral position inte längre blir möjlig.

En del av den kritik som lyfts rör ansvarsfrågor för effekter av automation och autonomt lärande processer och tjänster, låt vara troligen oväntade effekter. Ansvarsfrågor kan också adresseras för passivitet hos plattformar för sådant de genom sina automatiserade system kan upptäcka och som är av väldigt allvarlig karaktär – traffickingnätverk, vapensmuggling och självmordsbenägenhet⁵⁰. Vilken typ av passivitet är egentligen klandervärd? När AI-drivna och datainsamlade hemassistenter blir vardag i våra hem, vilket ansvar bör producenterna ha för sådant som de därmed upptäcker – eller skulle kunna upptäcka – i termer av olagligheter eller individens trängande behov av hjälp? När någon berättar om ett stundande självmord eller när barn misshandlas eller far illa på andra sätt? Med extensiv datainsamling av människors alla aktiviteter har ett samhälle vuxit fram som medievutare Mark Andrejevic kallat ett "sensorsamhälle",⁵¹ vilket av nödvändighet också väcker ansvarsfrågor för områden som tidigare inte adresserats. När plattformarna kan detektera och analytiskt förstå en rad allvarliga skeenden behöver de också ta ställning till den normativa frågan hur de ska agera – eller inte agera – gällande sina insikter.

49 "Should the data be representative of the world as it is, or of a world that many would aspire to?"; se även Larsson (2018c).

50 Se exempelvis Facebooks arbete med att upptäcka självmordsbenägenhet. TechCrunch (27 nov 2017) "Facebook rolls out AI to detect suicidal posts before they're reported", av Josh Constine. <https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/>

51 Andrejevic (2013).

Forskare vid det amerikanska institutet AI Now konstaterar att utan verkställbara mekanismer för transparens och revision av dataanvändningen eller fungerande ansvarsfördelning finns det knappast någon kännedom om hur algoritmiskt beslutsfattande begränsar eller hindrar vad de kallar medborgerliga rättigheter.⁵² Caplan et al. talar här om vikten av s.k. *algoritmiskt ansvarserkännande* (eller ”algorithmic accountability”), vilket har kommit att bli ett allt viktigare tema i den kritiska och samhällsinriktade AI- och ML-forskningen. Centralt för att ens kunna formulera ansvar eller lösningsinriktade förslag ligger därmed i hur man hanterar transparens, vilket är långt ifrån ett tydligt begrepp.

SJU NYANSER AV TRANSPARENS

För en rad algoritmberoende och konsumentinriktade digitala plattformar har vi redan system de allra flesta inte har insyn i, eller kunskap om hur besluten tas. Dock, eftersom ansvarsfrågorna är så nära förknippade med insyn och transparens – att konsument, företag och myndigheter ens kan vederlägga vad som sker inuti de i vart fall delvis automatiserade processerna – så är transparensfrågan helt central för att förstå algoritmers roll för digitala plattformar och de utmaningar som uppstår.⁵³ Kommissionen genomför en studie under 2018 och 2019 och analyserar s.k. *algoritmisk transparens* för att öka medvetenheten och bygga en god kunskapsgrund för utmaningar och möjligheter för algoritmiska beslut:

Algorithmic transparency has emerged as an important safeguard for accountability and fairness in decision-making and for opening

52 Caplan et al., (2018).

53 Vilket uttrycks både av EU-kommissionen (25 april 2018) såväl som i olika forskningsinriktningar, vilket refereras till under relevant terminologi nedan. För rättsliga exempel, se särskilt automatiserat beslutsfattande baserade på personuppgifter (GDPR) och högfrekvenshandel på aktiemarknaden (Markets in Financial Instruments Directive, MiFID II).

to scrutiny the way access to information is mediated online, especially on online platforms.⁵⁴

Det finns dock ett behov av att nyansera vad man avser med transparens när man uttrycker ett behov för det – för vem, på vilket sätt och i vilket syfte? För att kunna göra det så diskuterar jag sju nyanser av transparens nedan, som kommer av olika orsaker och relaterar till olika syften.⁵⁵ Detta betyder att jag här föreslår en vid och samhällsvetenskapligt viktad definition av de utmaningar som den s.k. ”svarta lådan” ger, som också är menad att förtydliga att bristen på transparens i många fall kan förklaras med att det finns flera motstående intressen som drar i andra riktningar. En utmaning, från ett samhällsperspektiv, ligger i att *balansera* motstående intressen, där punkt 1 och 2 nedan representerar motverkande transparensintressen och 3 till 7 utgör varianter av kunskaps- och andra insynsutmaningar.

1. Ägande: att mjukvara och data är *proprietär*, t.ex. ett (för transparensen motstående) företagsintresse att inte visa upp hur de har löst ett visst problem, vilket kan bli fallet när en produkt kommersialiseras och skalas upp på en marknad. För vissa företag är mjukvaran och dess algoritmer värdefulla ”recept” som de betraktar som företagshemligheter.⁵⁶
2. *Undvikande av missbruk* (”gaming”): att transparens skulle motverka processens syfte och därmed kan missbrukas eller manipuleras för att nå fördelar, som med trendingfunktionen hos Twitter, eller i fördelning av försörjningsstöd eller andra varianter av profilering eller betygssättande.⁵⁷
3. Kompetens: att det krävs en viss specifik kompetens för att ens kunna bedöma algoritmer och deras dataanvändning, vilket man

54 EU-kommissionen (25 april 2018). Algorithmic Awareness-Building. <https://ec.europa.eu/digital-single-market/en/algorithmic-awareness-building>

55 Se även Larsson (2019).

56 Spiekermann & Korunovska (2016).

57 Se exv. Caplan et al. (2018).

- kan kalla *datalitteracitet* eller *algoritm litteracitet*.⁵⁸
4. *Koncept, terminologier och förklarbarhet*: hur den komplexa AI-processen förklaras genom val av språk, metaforer och symboler, har direkta implikationer för hur den förstås, vilket i sin tur relaterar till acceptans och tillit.⁵⁹
 5. *Marknadskomplexitet*, vilket kan ses som en kombination av proprietära upplägg och att de datadrivna marknaderna kan ses som komplexa ”ekosystem” där data mäklaras och reser vidare mellan en rad aktörer. Detta inkluderar också det ofta kommersiellt motiverade upplägget med trackers, som tredjepartskakor och pixlar, dvs. det är ofta svårt att följa vart datan tar vägen.⁶⁰
 6. *Distribuerat, individualiserat utfall*, kan bli utfallet av konsumentprofilerande tjänster som vill ”personalisera” sin affär, sin prissättning eller marknadsföring – vilket rimligen utmanar inte minst tillsynsverksamhet.⁶¹
 7. *Algoritmkomplexitet*: att autonomi i den självlärande algoritmen är så självständig att ingen har egentlig insyn eller kunskap om hur den löser ett problem, bara att den löser det (eller inte). Utfallet kan vara en högre sannolikhet för ett visst utfall, som i en applicerad kontext skulle kunna omsättas i högre lönsamhet genom mer försäljning eller mer precision i en diagnos, men inte nödvändigtvis på vilka grunder eller hur resultatet nåddes i detalj.⁶² Tolkningsbar och *förklarbar* maskininlärning har varit ett fält i den inomvetenskapliga

58 Se även Jutta Haider och Olof Sundins bidrag i denna volym, och Haider & Sundin (2019) som utvecklar betydelsen av media- och informationslitteracitet i relation till sökmotorer.

59 Hur vi förstår abstrakta, exempelvis digitala, fenomen kan vara avgörande för både hur de regleras men också för hur vi normativt bedömer dem. För en omfattande genomgång av metaforer och konceptuella metaforers rättsliga betydelse i relation till digitala fenomen, se Larsson (2017a). Man kan också diskutera skillnaderna mellan tolkningsbarhet (”interpretability”), förklarbarhet (”explainability”) och förståbarhet (”comprehensibility”) för maskininlärningsmodeller i sig, se Guidotti et al. (2018).

60 Se Pasquale (2015) och Larsson (2018b). Komplexiteten i marknadsupplägget tecknas tydligt i Christl (2017).

61 Vilket jag utvecklar i förhållande till konsumentskydd i Larsson (2018a; 2018b). För en analys av individualisering i relation till public service, se Jockum Hildéns bidrag i denna volym.

62 Även AI-forskningen talar här om AI-explainability (AIX), se IEEE, 2018, jfr. Wachter et al. (2017).

datavetenskapliga litteraturen ett tag, ibland kallat AIX, men en kritisk genomgång visar på behovet av att klassificera problemen tydligare,⁶³ inte minst i relation till den ökade praktiska betydelsen,⁶⁴ och där även kunskap inom discipliner som socialpsykologi och kognitionsvetenskap skulle kunna bidra.⁶⁵

För Googles sökmotor, exempelvis, kan man tänka sig att det proprietära skälet (1) väger tungt – man vill inte visa sina konkurrenter exakt hur upplägget ser ut, och man vill också undvika missbruk (2) eftersom det skulle motverka det uppställda syftet om träffrelevans, som delvis har ett distribuerat, individualiserat utfall (6). Marknadskomplexitet (5) är rimligen en del av svårigheten för någon utomstående att teckna hur automatisering och interna policybeslut påverkar individer och företag, vilket även EU-kommissionens domar, nämnda ovan, vittnar om. Bredden och mängden av datainsamlade applikationer med miljarder användare – allt från sökmotorn och kartfunktionen till Chrome, Android, Gmail, Analytics, YouTube – gör också helheten svårbedömd utifrån.⁶⁶

Exempelvis myndigheternas tillsynsarbete kan ses som ett slags transparensarbete, där syftet är att granska vad marknadsaktörerna gör, för att därefter bedöma om det finns något otillbörligt i deras praktik. På konsumentområdet uppstår svårigheter för tillsynen när utfallet är distribuerat och individualiserat (6), exempelvis för digital marknadsföring och individuell prissättning. Detta talar för att tillsynsverksamheterna behöver utveckla datadrivna metoder för att upptäcka otillbörligheter i storskaliga plattformars tjänster, oavsett om de är avsiktliga eller ej.⁶⁷ Komplexiteten (5) och proprietärt motiverad (1) brist på

63 Guidotti et al. (2018).

64 Biran & Cotton (2017).

65 Miller (2019).

66 Analytics, exempelvis, används enligt en studie av omkring 75% av toppmiljonen mest populära sajter (se Englehardt & Narayanan, 2016). Samma studie konstaterar också att nyhetssajter har flest antal trackers överlag, troligen i syfte att skapa reklamintäkter som plats för riktade annonser, men även som målgruppsanalys. Se även Larsson 2017b för mer om bristen på transparens i s.k. programmatisk reklam.

67 Vilket jag utvecklat på annan plats i termer av ”algorithmic governance”, Larsson (2018a).

insyn lyfts av Frank Pasquale som en problematisk ”black box” för digitala marknader.⁶⁸

Bristen på transparens blir påtaglig när plattformar missbrukar sin informationsställning. The New York Times avslöjade exempelvis i mars 2017 att Uber hade utvecklat en hemlig operation som kallades Greyball och som inkluderade mjukvara för att undvika polisiär utredning och granskning av deras tjänst.⁶⁹ Mjukvaruverktyget identifierade vissa passagerare som troliga poliser eller utredare och nekade dessa Uberbilar. De använde bl.a. ett slags geofencing runt myndighetsbyggnader och flaggade passagerare runt dessa byggnader som möjliga poliser. I Stockholm avslöjades detta vara fallet runt Polishuset på Kungsholmen. Uber blockerade även användare med mailadresser som slutade på polisen.se.⁷⁰ Uber hade en fejkversion av appen som flaggade personer försågs med, som kunde visa ”spökbilar” och göra Uberbilar otillgängliga.

Att AI-system bör vara förklarbara (4+7) poängteras både inom AI-forskningen såväl som i policydiskussioner, exv. i relation till GDPR och automatiserat beslutsfattande.⁷¹ Förklaringar kan dock ske på olika nivåer, med olika typ av symbolik och med olika typer av sociala behov.⁷² Vi är rimligen redan idag alla användare av en rad teknologier som vi bara har en ytterst vag förståelse för hur de egentligen fungerar. En ”förklaring” av exempelvis AI-genererat beslutsfattande kommer därmed oundvikligen att behöva välja nivå av konkretion genom val av symboler eller metaforer. Hur förklaringen kring AI-genererade processer ser ut i termer av metaforer och symbolik kommer därmed rimligen att vara väldigt viktigt för hur de kommer att uppfattas eller accepteras.

68 Pasquale (2015).

69 The New York Times (3 mars 2017) ”How Uber Deceives the Authorities Worldwide” av Mike Isaac. <https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html>; se även Rosenblat, 2018.

70 DiDigital (10 oktober 2017) ”Uber svartlistade svenska poliser – spårades via appen”, av Sven Carlsson. <https://digital.di.se/artikel/uber-svartlistade-svenska-poliser-sparades-via-appen>

71 Exv. i EU-kommissionens kommuniké om artificiell intelligens: ”...to further strengthen trust, people also need to understand how the technology works, hence the importance of research into the explainability of AI systems” (COM (2018) 237 final, s. 14).

72 Doshi-Velez (et al. 2017).

Det kan finnas starka konkurrensskäl för företag att inte visa upp hur deras mjukvara ser ut, men också för att undvika missbruk som motverkar syftet med en tjänst (2). Detta kan gälla försäkringsföretags bedömningar, såväl som Googles relevansbedömning i sökmotorn som Försäkringskassans val av variabler för hantering av vem som har rätt till ersättning.⁷³ Caplan et al. (2018) påpekar exempelvis att även den minsta öppenheten kring hur trendingfunktionen på Twitter fungerar har gjort det möjligt att manipulera delar av miljön för att täcka vissa ämnen med automatiserade bottar eller bot-nätverk för att påverka och manipulera debatter. Omvänt kan det finnas extra starka skäl till transparens hos myndigheters användning av automatiserat beslutsfattande, där rättssäkerhet och likabehandling är centrala för beslutsfattandet som därmed riskerar att bli avhängigt hur algoritmer designas av upphandlade företag.⁷⁴

Beroende på vad man menar med transparens och beroende på vilken av de sju nyanserna man vill utveckla eller förlita sig på så kommer olika typer av åtgärder att svara upp mot respektive kategori. Om problem-bilden adresserar det (1) proprietära upplägget så kommer argumentationen få göras avhängig att samhällsintresset är större än företagets behov av hemligheter, vilket möjligen kunde undergräva konkurrens. En kompromiss som finns inom andra sektorer vore att lyfta vad Pasquale (2015) kallat en ”kvalitativ transparens” där en dedikerad part kan beredas insyn, men inte att en generell öppenhet måste råda.⁷⁵ Om problem-bilden relaterar till den (7) algoritmiska komplexiteten och brist på insyn och förklarbarhet som vissa AI-processer kan erbjuda så är rimligen forskarsvaret att försöka minska på denna brist, och policysvaret antingen att reglera eller utveckla standarder i samspel med marknadsaktörer där de sektorer som allra mest behöver transparens inte tillåter de mest icke-transparenta verktygen. Vilket också kommer att behöva

73 Jfr. Larsson (2019); Caplan et al., (2018).

74 Jfr. Svensson och Larsson (2018). Se Reisman et al. (2018) som föreslår en ”algorithmic impact assessment” för användning av automatiserade algoritmiskt styrda processer i offentlig verksamhet.

75 Jfr. Larsson (2018a).

inkludera (4) konceptuell och terminologisk utveckling. Forskare har även pekat på behovet av ”auditability”, dvs. granskningsmöjlighet, för tredje parter att sondera och granska beteendet hos en algoritm.⁷⁶

Kompetensfrågan (3) har identifierats av både regering och innovationsmyndigheter för närvarande, men då mer som en innovationsfråga snarare än en granskningsfråga. Från ett tillsynsperspektiv är rimligen kompetensfrågan avgörande, precis som (5) marknadskomplexiteten och det (6) distribuerade utfallet. Avslutningsvis bör man komma ihåg att inte all brist på transparens motverkar processens syfte. För vissa processer vill man helt enkelt undvika missbruk (2) av hur processen är tänkt att fungera.

SLUTSATSER

Transparensfrågan kring AI-drivna processer är en central fråga som träffar digitala plattformar, precis som Mark Zuckerberg konstaterade i de amerikanska senatsförhören i april 2018, indikerat i kapitlets inledande citat. Olika sammanhang kan dock kräva olika nivåer av insyn, beroende på vad syftet och behovet är. För vissa fall skulle en dedikerad eller kvalitativ transparens vara att föredra, likt en bilbesiktning där inte var och en konsument egentligen har möjlighet eller ens behov av att granska innehållet och funktionerna, men det samhälleliga intresset av att utfallet av en viss applikation inte är otillbörligt och diskriminerande beaktas. För andra fall skulle en radikal öppenhet kunna vara eftersträfvansvärd utifrån ett samhällsperspektiv. Vad man avser med transparens och begrepp som ”förklara” och att ”förstå” kommer rimligen också att behöva förtydligas ytterligare för var och en kontext där algoritmiskt beslutsfattande appliceras, speciellt i förhållande till ansvarsfrågor och s.k. *algorithmic accountability*. Storskaligheten hos många digitala plattformar och den datadrivna automatiseringen leder

76 Diakopoulos & Friedler (2016).

till att den samhälleliga betydelsen och därmed förtydligandet av riskbilden blivit oerhört viktig. Oförutsedda konsekvenser av plattformarnas normativa arbete med automatiserad policyimplementering har kommit att bli helt centrala att förstå för de marknadsskapande och infrastrukturella upplägg som påverkar hela marknader, inklusive individer och företag.

Att argumentera för att de digitala plattformarna är normativa och innehar rättsliga attribut och därmed bär ansvar för denna normativitet är dock inte nödvändigtvis samma sak som att kräva transparens i allt de gör. Det finns, som visat ovan, flera goda skäl till att vissa processer inte bör vara transparenta. Det innebär heller inte att de aldrig kan trampa snett eller ta enskilda dåliga beslut. Det innebär däremot att den viktiga tilliten till plattformsaktörerna är beroende av att de artikulerar motiven bakom sitt policyarbete, visar på de konsekvenser de eftersträvar och därmed accepterar sin samhälleliga betydelse. Precis som rättssystem är beroende av att ständigt stresstestas och utvärderas gentemot ett föränderligt samhälle – kulturella normer förändras över tid, ny teknik katalyserar förändringar i förutsättningar för hela branscher. Reglering av plattformsaktörer är inte heller nödvändigtvis en utvecklingsbromsande faktor, som en relativt vanlig men ofta förenklad syn vill hävda. Tvärtom, de samhälleliga behoven kan ibland bättre säkerställas, inklusive tilliten till både kommersiell dataanvändning och automatiserade verktyg, om spelreglerna är tydliga för alla inblandade. Den samhälleliga frågan handlar också påfallande ofta om att balansera legitima men delvis oförenliga intressen.

Maskininlärning och andra varianter av artificiell intelligens som används som central part i applikationer som interagerar med människor och samhälleliga strukturer samspekar därmed ofrånkomligen med historiska upplägg och samhällelig bias. I många fall finns därmed ingen normativt neutral mark: ska samhällelig bias reproduceras eller motverkas av den specifika applikationen? Ska skevheter i genus, makt, etnicitet, ekonomi och religion motverkas eller utgöra underlag

för relevansbedömningar, prissättning och andra åtgärder? Och, om den ska motverkas, måste även dessa normer definieras – vilka normer ska styra? Hur ska balansen definieras? Att införliva mer kompetens om kulturer, normer och etik i utvecklingen av digitaliserad automation bör inte bara motiveras som ett konstruerat efterhandspålägg som bromsar viss användning av AI och maskininlärning. Denna typ av kompetens, som också behöver införlivas i det AI-relaterade kursutbudet vid tekniska fakulteter, kan möjligen också bidra till bättre applikationer – dvs. som en kärnfråga på designsidan, demonstrerat av några av de undermåliga applikationer och värderingsgrundande automatiseringsmisstag redovisade ovan.⁷⁷

Man kan konstatera att digital plattformisering innebär en allt viktigare samhällelig organisationsform som politiker, forskare och näringslivsföreträdare behöver förstå implikationerna av, i takt med att fler verksamheter anammar en datadriven, skalbar och automatiserad praktik. Plattformar skapar och genomför policyer som dagligen sammanför – påverkar, reglerar – miljarder vardagsliv och företags förutsättningar i detalj. De är, i vissa fall, globala och därmed i någon mening supra-jurisdiktionella, dvs. svårreglerade genom traditionell nationell reglering. Även om de globala plattformarna ofta strävar efter generellt överskridande policies så nödgas de ofrånkomligen i allt högre utsträckning att formulera och implementera mer regionala, nationella och lokala hänsyn. GDPR är ett tydligt exempel gällande dataskydd för EU kontra omvärlden, men policy gäller inte bara jurisdiktioner – som för övrigt kan innehålla direkt motstridig reglering olika rättsordningar emellan – utan även kulturella kontexter och sociala normer hos olika grupper världen över men också inom samma samhälle.

Vidare är det organisatoriska policybestämmandet inom plattformarna i hög grad centraliserat, med en i hög grad automatiserad implementering. Det finns en betydande samhällsutmaning i att de digitala plattformarna därmed utgår från delvis dolda policies som implementeras

77 Se exv. p. 10 i AI Now Report 2018, Whittaker et al.

genom i hög grad dolda förfaranden och att de känsligaste fallen avgörs av i hög grad dolda domstolar. Klart är dock att de digitala plattformarna och en växande plattformisering av såväl privata som offentliga verksamheter är här. Artificiell intelligens och maskininlärning representerar ett löfte i linje med denna del av digitaliseringen, samtidigt som de matematikdrivna teknologierna väcker nya och fundamentalt viktiga samhällsfrågor i termer av ansvarserkännande och ansvarsfördelning, vilken nivå och typ av transparens som är lämplig för vad, och ytterst hur tillit byggs och relaterar till autonoma och artificiella beslut. Transparensen hos samhällsapplicerad AI behöver bli högre, av flera skäl. Några av dess utmaningar i förhållande till de digitala plattformarnas automatiserade policyimplementering har tecknats i detta kapitel. Behovet relaterar förvisso till frågan om hur plattformar bäst bör regleras, men kanske ännu mer till att bättre förstå hur plattformarna de facto reglerar oss. ■

REFERENSER

Andersson Schwarz, J. (2017). Platform Logic: An Interdisciplinary Approach to the Platform-Based Economy. *Policy & Internet* 9(4): 374–394.

American Scientific (20 juni 2016). "When Will Computers Have Common Sense? Ask Facebook", av Larry Greenemeier. <https://www.scientificamerican.com/article/when-will-computers-have-common-sense-ask-facebook/>

Andersson Schwarz, J. & Larsson, S. (2018). "A Platform Society", i Larsson & Andersson Schwarz (red.) *Developing Platform Economies. A European Policy Landscape*. Brussels and Stockholm: European Liberal Forum asbl and Fores.

Andrejevic, M. (2013). *Infoglut. How too Much Information is Changing the Way We Think and Know*. New York, NY: Routledge.

Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. In *IJCAI-17 Workshop on Explainable AI (XAI)*.

Business Insider (13 jan 2018). "Facebook is trying to prove it's not a media company by dropping the guillotine on a bunch of media companies" av Steve Kovach. <https://www.businessinsider.com/facebooks-updated-news-feed-algorithm-nightmare-for-publishers-2018-1?r=US&IR=T&IR=T>

Buzzfeed (11 april 2018). "Why Facebook Will Never Fully Solve Its Problems With AI" av Davey Alba. <https://www.buzzfeednews.com/article/daveyalba/mark-zuckerberg-artificial-intelligence-facebook-content-pro>

Caplan, R., Donovan, J., Hanson, L. & Matthews, J. (2018). *Algorithmic Accountability: A Primer*, NYC: Data & Society.

Christl, W. (2017). *Corporate Surveillance in Everyday Life: How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions*. Vienna: Cracked Labs.

Diakopoulos, N., & Friedler, S. (2016). How to hold algorithms accountable. *MIT Technology Review*, 17(11).

DiDigital (10 okt 2017). "Uber svartlistade svenska poliser – spårades via appen", av Sven Carlsson. <https://digital.di.se/artikel/uber-svartlistade-svenska-poliser-sparades-via-appen>

Doshi-Velez, F., Korts, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Shieber, S., Waldo, J., Weinberger, D. & Wood, A. (2017). Accountability of AI under the law: The role of explanation. *arXiv preprint arXiv:1711.01134*.

Englehardt, S. & Narayanan, A. (2016). "Online Tracking: A 1-million-site Measurement and Analysis," extended version of paper at ACM CCS 2016.

EU-kommissionen (25 april 2018). "A European approach on Artificial Intelligence", Bryssel.

EU-kommissionen (25 april 2018). Algorithmic Awareness-Building, <https://ec.europa.eu/digital-single-market/en/algorithmic-awareness-building>

EU-kommissionen (27 juni 2017) Press release. "Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service". Bryssel. http://europa.eu/rapid/press-release_IP-17-1784_en.htm

EU-kommissionen (18 juli 2018) Press release. "Antitrust: Commission fines Google €4.34 billion for illegal practices regarding Android mobile devices to strengthen dominance of Google's search engine". Bryssel. http://europa.eu/rapid/press-release_IP-18-4581_en.htm

EU-kommissionen (25 april 2018) Artificial Intelligence for Europe (COM(2018) 237 final).

Facebook Transparency Report: <https://transparency.facebook.com>

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T. (2018b). "The Logan Paul YouTube controversy and what we should expect from internet platforms. When users abuse a platform, who is responsible?" Vox, 16 jan. <https://www.vox.com/the-big-idea/2018/1/12/16881046/logan-paul-youtube-controversy-internet-companies>

The Guardian (12 jan 2018). "Google's solution to accidental algorithmic racism: ban gorillas", av Alex Hern. <https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people>

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5): 93.

Haider, J., & Sundin, O. (2019). *Invisible Search and Online Search Engines: The ubiquity of search in everyday life*. Chicago: Routledge Studies in Library and Information Science.

The Intercept (29 nov 2018). "Google Shut out Privacy and Security Teams from Secret China Project", av Ryan Gallagher. <https://theintercept.com/2018/11/29/google-china-censored-search/>

IEEE (2018). *Ethically Aligned Design. A vision for prioritizing human well-being with autonomous and intelligent systems*. Version 2.

Larsson, S. (2019). "Artificiell intelligens som normativ samhällsutmaning: partiskhet, ansvar och transparens", Lund: Juristförlaget.

Larsson, S. (2018a). Algorithmic Governance and the Need for Consumer Empowerment in Data-driven Markets, *Internet Policy Review* 7(2):1-12.

Larsson, S. (2018b). "Den kvantifierade konsumenten: Om behovet av tillit och transparens på datadrivna marknader". *Policy brief 2018:1*. Stockholm: Fores.

Larsson, S. (2018c). "Sjysta Algoritmer och normativ design", i Akenine, D. & Stier, J. (red.) *AI och människor*. Stockholm: AddAI.

Larsson, S. (2017a). *Conceptions in the Code. How Metaphors Explain Legal Challenges in Digital Times*. Oxford University Press.

Larsson, S. (2017b). "Digital konsumentprofilering. Stora data, prediktiv analys och policyutmaningar", i Sandberg, A. (red.) *Kunskapsöversikter inom det konsumentpolitiska området*, Konsumentverket. 2017:1.

Larsson, S. & Ledendal, J. (2017). *Personuppgifter som betalningsmedel*, Karlstad: Konsumentverket. 2017:4.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. Vol 267: 1-38. <https://doi.org/10.1016/j.artint.2018.07.007>.

The New York Times (3 mars 2017). "How Uber Deceives the Authorities Worldwide" av Mike Isaac. <https://www.nytimes.com/2017/03/03/technology/uber-grey-ball-program-evade-authorities.html>

The New York Times (18 nov 2018). "China Rules. How China became a superpower", av Raymond Zhong. <https://www.nytimes.com/interactive/2018/11/18/world/asia/china-internet.html>

The New York Times (20 okt 2018) Debatt: "No, A.I. Won't Solve the Fake News Problem. At least not for decades to come. Sorry, Mark Zuckerberg." Av Gary Marcus och Ernest Davis. <https://www.nytimes.com/2018/10/20/opinion/sunday/ai-fake-news-disinformation-campaigns.html>

Noble, S. U. (2018). *Algorithms of Oppression: How search engines reinforce racism*. NYU Press.

Pasquale, F. (2015). *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press.

Pichai, S. (18 juli 2018). "Android has created more choice, not less", Google in Europe. <https://www.blog.google/around-the-globe/google-europe/android-has-created-more-choice-not-less/>

ProPublica (14 sep 2017). "Facebook Enabled Advertisers to Reach 'Jew Haters'" av Julia Angwin, Madeleine Varner and Ariana Tobin. <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>

Radiolab (17 aug 2018). "Post no evil". <https://www.wnycstudios.org/story/post-no-evil>

Regeringskansliet (2018). *Nationell inriktning för artificiell intelligens*. Näringsdepartementet.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic Impact Assessment. A practical Framework for Public Agency Accountability*. NYC: AI Now.

Rosenblat, A. (2018). *Uberland: How Algorithms Are Rewriting the Rules of Work*. University of California Press.

Spiekermann, S., & Korunovska, J. (2016). Towards a value theory for personal data. *Journal of Information Technology*, 23(1): 62-84. doi:10.1057/jit.2016.4

Svensson, L. & Larsson, S. (2018). *Digitalisering av kommunal socialtjänst. En empirisk studie av en organisation och profession i förändring*. FoU-rapport 2018:1. Helsingborg Stad.

TechCrunch (12 jan 2018). "Facebook feed change sacrifices time spent and news outlets for 'well-being'" av Josh Constine. <https://techcrunch.com/2018/01/11/facebook-time-well-spent/>

TechCrunch (27 nov 2017). "Facebook rolls out AI to detect suicidal posts before they're reported", av Josh Constine. <https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/>

The Telegraph (29 nov 2018). "Apple removes over 700 apps from Chinese App Store" av James Titcomb. <https://www.telegraph.co.uk/technology/2018/11/29/apple-removes-700-apps-chinese-app-store/>

Tufekci, Z. (10 mars 2018). "YouTube, the Great Radicalizer", NY Times opinion. <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html?referer=>

Vaidhyanathan, S. (2017). "Facebook Wins, Democracy Loses". Op-ed i *The New York Times*, 8 sept. <https://www.nytimes.com/2017/09/08/opinion/facebook-wins-democracy-loses.html>

van Dijck, J., Poell, T., & de Waal, M. (2018). *The Platform Society. Public Values in a Connective World*. Oxford University Press.

Wachter, S., Mittelstadt, B. & Floridi, L. (2017). "Transparent, Explainable, and Accountable AI for Robotics." *Science Robotics* 2, no. 6 (2017): eaan6080. doi:10.1126/scirobotics. aan6080

Whittaker, M. Crawford, K., Dobbe, R., Fried, G., Kazianus, E., Mathur, V., Myers West, S., Richardson, R., Schultz, J. & Schwartz, O. (2018) *AI Now Report 2018*. New York: NYU.

The Washington Post (10 april 2018). "Transcript of Mark Zuckerberg's Senate hearing", https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?noredirect=on&utm_term=.c9de4ee99abc

Zou, J. & Schiebinger, L. (18 juli 2018). "AI can be sexist and racist — it's time to make it fair", *Nature*, comment.

STEFAN LARSSON är docent i teknik och social förändring vid institutionen för teknik och samhälle på LTH, Lunds universitet och programchef för det digitala samhällsprogrammet på tankesmedjan Fores. Han är jurist och disputerad i såväl rättssociologi som i fysisk planering. Hans forskning är inriktad mot digitalisering, datadrivna marknader och AI, med fokus på tillit, transparens och ansvar.



Detta är ett särtryck (2018) från antologin *Plattformssamhället. Den digitala utvecklingens politik, innovation och reglering* (Andersson Schwarz och Larsson, red. 2019), utgivet av tankesmedjan Fores under Creative Commonslicensen Erkännande-Ickekommersiell-IngaBearbetningar 4.0 Internationell (CC BY-NC-ND 4.0).

Sprid gärna!