

Spaningar från Silicon Valley

The Blind Spots

Algoritmernas blinda fläckar – när
AI-revolutionen glömmer halva mänskligheten



Om serien ”Spaningar från Silicon Valley”

Den artificiella intelligensen har på kort tid revolutionerat våra liv, men digitaliseringen av vårt samhälle har pågått under en mycket längre tid än så. De kvinnor som har haft oundgängliga roller för digitaliseringen genom historien får inte den uppmärksamhet de förtjänar för sina insatser och i dag är det män som utvecklar och driver de radikala innovationerna inom exempelvis artificiell intelligens. För att förstå vilka problem det leder till och hur vi kan se till att undvika dem i framtiden, tog jag initiativet till antologin ”Felkod kvinna”.

Futurion har också uppmärksammat hur AI förändrar spelplanen för unga. Det handlar inte bara om teknik, utan om vem som får chansen att komma in, bygga erfarenhet och börja ett arbetsliv. Forskning från Örebro universitet bekräftar nu att unga tappar mark i AI-exponerade yrken, medan äldre gynnas. Samtidigt visar Futurions mätningar att många unga inte tror att politiken riktigt fattar vad som håller på att hända.

Tillsammans behöver vi ändra på det. Ett sätt är att spana med hjälp av journalisten Katarina Andersson, som är baserad i San Francisco – mitt i miljön där AI-utvecklingen, arbetslivstrenderna och de kulturella skiftena faktiskt sker.

Ann-Therése Enarsson, vd, Futurion



Om författaren



Katarina Andersson är frilansjournalist och har bevakat digitaliseringsfrågor i många år. Hon har bland annat jobbat som techkorrespondent för Aftonbladet i San Francisco, producerat radioserier om samhällets digitalisering för Sveriges Radio och frontat techsajten Breakits podd och liveshow.

Grafisk formgivning: Narva Communications

Omslag: AI-genererad bild

Tryck: PÅ Media

ISBN 978-91-989527-6-6

Inledning

AI beskrivs ofta som ett tekniskt problem: bättre data in, bättre beslut ut. Men i praktiken handlar det om något helt annat. När AI nu börjar styra allt från rekrytering till produktutveckling växer en obekväm insikt fram: systemen speglar inte bara världen som den är – de riskerar att förstärka vem som får plats i den. Och i vissa fall avgör de också vem som inte får det.

Att AI-system kan bli skeva är inget nytt. Forskning från bland annat Stanford och MIT har gång på gång visat samma sak: modellerna lär sig av den data – och de människor – som bygger dem. Och när både datan och teamen är snävt sammansatta, blir resultaten därefter. Det här är inte en teoretisk risk – utan något som redan fått konsekvenser.

I den uppmärksammade "Gender Shades"-studien testades ansiktsigenkänningsystem från flera stora techbolag. Resultatet var slående: felmarginalen var upp till 34 procent för mörkhyade kvinnor – jämfört med under 1 procent för ljushyade män. Sådana brister har i verkligheten lett till att människor pekats ut som misstänkta. I februari 2023 blev Porcha Woodruff, då höggravid i åttonde månaden, gripen av polis i sitt hem i Detroit. Anklagelsen gällde ett rån hon inte begått; algoritmen hade trampat helt fel. Fallet blev snabbt ett skräckexempel. Samma mönster går igen på arbetsmarknaden. Amazons rekryteringsverktyg fick skrotas efter att den systematiskt sorterat bort kvinnliga kandidater, helt enkelt eftersom modellen tränats på historisk data från en mansdominerad bransch.

Samtidigt visar forskning från MIT att team med större mångfald inte bara är mer rättvisa – utan också bygger mer träffsäkra och robusta system. Fler perspektiv gör att blindade fläckar fångas upp innan de hinner bli problem. Men, i praktiken är sådana team fortfarande undantag, inte regel.

Kort sagt: skev AI är inte en hypotetisk risk. Den är redan här.



Gender Shades-studien från MIT Media Lab

DENNA MILSTOLPE visade att felmarginalen för mörkhyade kvinnor var så hög som 34,7 %, medan den för ljushyade män var nästintill obefintlig (0,8 %). Resultatet ledde till att flera stora tech-bolag tvingades se över sina algoritmer och i vissa fall pausa försäljningen av ansiktsgenkännings-teknik till polismyndigheter.

Arbetsmarknadens obalans

ENLIGT RAPPORTER från bland annat IMF löper kvinnor 11 % högre risk än män att få sina jobb automatiserade av AI, främst på grund av överrepresentation i administrativa roller.

Historisk data

AMAZONS AI tränades på CV:n inskickade under en 10-årsperiod. Målet var att bygga den "heliga graalen" inom rekrytering: en algoritm som kunde scanna 100 CV:n och spotta ut de fem bästa. Men eftersom modellen tränades på historisk data från en mansdominerad bransch drog den slutsatsen att framgång var synonymt med att vara man. Det ledde till att ordet "kvinnors" (som i "kvinnors schackklubb") automatiskt sänkte kandidatens betyg. 2017 tvingades Amazon skrota projektet helt. Trots försök lyckades man inte tvätta bort fördomarna. Det blev ett bevis på att AI som tränats på orättvis dåtid aldrig kan bygga en rättvis framtid.

Varför AI blir skev

AI-modeller tränas på enorma mängder data – texter, bilder, beteenden – som människor har skapat. Det betyder att systemen inte bara lär sig fakta, utan också våra mönster: vad vi klickar på, vilka vi anställer, hur vi uttrycker oss.

Problemet är att den datan ofta är sned från början. Om en bransch historiskt dominerats av män, eller om vissa grupper syns mindre i data, lär sig modellen att det är "normalt". Och när systemen sedan optimeras för det som är vanligast – det majoriteten gör – förstärks mönstret.

När AI hittar det vi missar

AI kritiseras ofta för att förstärka bias och cementera gamla mönster. Men det finns också ett annat perspektiv: rätt använd kan tekniken göra det motsatta och synliggöra det som är skevt. Det är utgångspunkten för Yana Welinder, en av de mest intressanta svenskarna på Silicon Valleys AI-scen – men nästan okänd på hemmaplan.

Förra året sålde hon sin AI-startup Kraftful till produktanalysjätten Amplitude. Hennes idé sticker ut: AI kan inte bara förstärka bias – utan också avslöja den. Med Kraftful, baserat i San Francisco, visade Yana Welinder tidigt att AI inte bara handlar om att chatta – utan om att fatta bättre affärsbeslut. Amplitude är världsledande på produktanalys. De ser vad användare gör – klickar på knappar, stänger ner appen eller köper saker. Men de saknade en viktig pusselbit: Varför användarna gör som de gör.

Och det var den pusselbiten som Welinders Kraftful kunde leverera. AI-systemet läser recensionerna och lyssnar på användarfeedback. För henne handlar frågan om mångfald inte bara om vilka som bygger AI – utan om hur systemen faktiskt lyssnar. I ett av hennes tidigare projekt analyserade de stora mängder användardata från dejtingappar. Målet var enkelt: förstå vad användarna egentligen ville ha. Men ett mönster stack ut.

– Det var ett av de vanligaste klagomålen – och det lät först ganska obehagligt. Nästan rasistiskt, säger hon. Vad hennes team kunde konstatera var att många användare ville matchas med personer av samma etnicitet. Det här var i kölvattnet av George Floyd-protesterna och Black Lives Matter-rörelsens peak. Att ens ifrågasätta matchningsresultaten ur den vinkeln var känsligt.

Teamet tvekade. Men i stället för att avfärda signalen valde de att gräva vidare. Och när man tittade närmare på datan visade sig också att det som teamet förväntat sig – att vita ville matchas med vita – inte stämde.

– Det var faktiskt tvärtom. Det var minoriteter som klagade på att de fick dåliga matcher – för att de fick så få träffar på personer med samma etnicitet.



Yana Welinder.

”När man optimerar på vad de flesta gör så suddas minoriteterna ut.”

Och Welinders team kunde också förstå varför. De flesta dejtingappar var optimerade på beteendedata: vad folk klickar på, hur de swipar, vilka matchningar som sker. Och i den typen av data vinner majoriteten automatiskt.

– När man optimerar på vad de flesta gör så suddas minoriteterna ut. Resultatet: majoriteten fick bättre träffsäkerhet. Minoriteter fick sämre – utan att någon riktigt såg det. Det var alltså först när Welinders team använde AI för att analysera vad användare faktiskt skrev – i recensioner, feedback och kommentarer – som mönstret blev tydligt.

– När man faktiskt lyssnar på vad alla säger så kan man dra fram minoritetsröster med AI, säger Welinder. Här blir AI något annat än den ofta kritiserar för. I stället för att förstärka bias – kan den avslöja den. Skillnaden ligger i hur den används. Traditionell analys tittar på snittet. AI kan läsa allt.

– Vi behöver inte längre missa nyanser, säger Welinder. Och just där ligger den verkliga vändningen. Problemet var aldrig att datan saknades. Problemet var att ingen kunde se den. När AI används för att faktiskt

lyssna – inte bara optimera – blir den ett verktyg för att upptäcka mönster som tidigare varit dolda. Eller det som försvunnit i majoritetens brus. Det gäller inte bara dejtingappar. Samma logik finns i allt från offentliga tjänster till politiska plattformar – överallt där beslut fattas med data som grund.

– Om man använder den rätt kan AI avslöja bias i stället för tvärtom, säger Welinder.



Noreena Hertz:

”AI slår olika – och vi är inte redo”

Länge har AI-debatten handlat om modellerna: hur de tränas, vilka bias de innehåller och hur de kan göras mer rättvisa. Men enligt ekonomen Noreena Hertz är det bara början. Den verkliga frågan är vad som händer när tekniken lämnar labben – och börjar omforma arbetsmarknaden, ekonomin och människors liv. Hertz har gjort sig känd för att kombinera statistik, psykologi och hundratals djupintervjuer för att förstå hur samhällsförändringar faktiskt slår mot människor. När hon nu analyserar AI-revolutionen är hennes slutsats tydlig: effekterna kommer inte att slå jämnt.

– AI kan spela ut väldigt olika beroende på kön, bakgrund och etnicitet. Det är något vi inte riktigt har vaknat upp till ännu, säger hon till Futurion. För Hertz är frågan därför inte om tekniken är neutral – utan vem som sätter agendan för hur den utvecklas och vilka konsekvenser man är beredd att ta ansvar för.

Myten om den neutrala tekniken

Att AI-modeller tränade på historisk data reproducerar skevheter är i dag välkänt. Ett ofta citerat exempel är Amazons rekryteringsverktyg, som nämnts i inledningen av den här rapporten. Verkttyget fick skrotas efter att det systematiskt nedvärderat kvinnliga kandidater. Men enligt Hertz är problemet större än enskilda misstag.

– Att bryta ner köns- och etnisk bias hamnar inte högst på listan, säger hon. Den ofta upprepade föreställningen i techbranschen – att tekniken i sig är neutral – är enligt henne en myt. I stället måste systemen ständigt granskas: vilken data de tränats på, i vilken kontext – och med vilka prioriteringar. Hertz tvekar inte att utmana techledare offentligt. Hon lyfter Elon Musks AI-modell Grok som ett varnande exempel.



Noreena Hertz

Brittisk ekonom, akademiker, författare.

Doktor i ekonomi från Cambridge och en MBA från Wharton. En av världens mest inflytelserika röster inom ekonomi, samhällsutveckling och AI. Hedersprofessor vid University College London och sitter i styrelser för bolag som Mattel, Warner Music Group och Workhuman. Bästsäljande författare av böcker som *The Silent Takeover*, *The Debt Threat*, *Eyes Wide Open* och *The Lonely Century*.

– Jag drar mig inte för att säga “shape up”, säger hon. Modellen har kritiserats för att reproducera sexistiska stereotyper och för att släppa igenom antisemitiskt innehåll. När Musk beskriver Grok som “sanningssökande” menar Hertz att det speglar en bredare idé i delar av techvärlden: att systemen ska vara mindre filtrerade och mer “råa”. Men i själva verket tränas de på material där gamla fördomar fortfarande finns kvar – och därför också riskerar att förstärkas.

En arbetsmarknad som slår snett

Men för Hertz handlar AI inte bara om hur modellerna fungerar. Det handlar om hur hela arbetsmarknaden förändras – och vem som drabbas. Här ser hon ett tydligt mönster.

– Effekterna kommer att slå oproportionerligt mot kvinnor, som riskerar en dubbel, eller till och med trippel, smäll, säger hon. Kvinnor är överrepresenterade i yrken som administrativa assistenter, kundtjänst och kommunikation – precis de områden där AI nu snabbt kan effektivisera eller ersätta arbetsuppgifter.

Samtidigt är kvinnor underrepresenterade i de mest skyddade positionerna högre upp i hierarkierna. Och mittensiktet – där många kvinnor finns – riskerar att urholkas när AI gör det möjligt att “göra mer med färre”. Hertz drar paralleller till 1980-talet, när datoriseringen slog ut många kvinnodominerade kontorsjobb. Återhämtningen tog lång tid – och skedde ofta till lägre löner. Hon ser tecken på att något liknande kan vara på väg att hända igen. Dessutom pekar hon på en ny risk: att kvinnor både får mindre AI-utbildning och använder tekniken mindre i arbetslivet – vilket kan bli en nackdel när “AI-kompetens” blir ett krav.

Vad krävs nu?

För Hertz är det avgörande att utvecklingen inte lämnas åt marknaden ensam.

– Det är helt avgörande att regeringar och fackförbund vaknar – så att vi inte sömngångaraktigt kliver rakt in i en potentiell katastrof, säger hon. Hon pekar ut flera konkreta områden där politik och fack behöver agera.

Transparens och ansvar

När AI används i rekrytering eller uppsägningar måste det finnas rätt att förstå och ifrågasätta besluten – och få möjlighet till mänsklig prövning.

Vem ska betala för omställningen?

När AI tar över arbetsuppgifter behöver kostnader för omställningen fördelas. Här lyfter hon idén om en form av “robotskatt”, där företag som ersätter människor med AI bidrar – till exempel genom vidareutbildning, stöd till personer som förlorar jobb och investeringar i nya typer av arbeten.

”Det handlar inte bara om vad AI kan göra, utan om vem som tar ansvar för konsekvenserna”

Nya ägarmodeller

Ett mer radikalt förslag är att staten tar en liten ägarandel i AI-bolag och samlar den i en fond – en slags AI-variant av en statlig förmögenhetsfond – för att finansiera de samhällskostnader tekniken skapar.

– Det handlar inte bara om vad AI kan göra, utan om vem som tar ansvar för konsekvenserna – och vem som betalar när människor slås ut, säger Hertz.





Slutsats: Från passiva användare till Data Rewirers

Den stora utmaningen med AI är inte tekniken i sig, utan vår föreställning om att den är en naturkraft – något som bara händer oss. Men algoritmer är inte väderleksrapporter. De är arkitektur. Och precis som hus kan byggas med för smala dörrar eller trappor som ingen kan gå i, kan digitala system byggas så att de exkluderar.

De vi kallar för Data Rewirers är de som vägrar acceptera de befintliga kopplingarna. Det är produktutvecklarna som Yana Welinder, som använder tekniken för att lyssna på rösterna i marginalen. Det är ekonomerna som Noreena Hertz, som kräver politiskt ansvar för de sociala kostnaderna. Och det är de fackförbund och beslutsfattare som inser att transparens är en grundförutsättning för en fungerande demokrati.

Att ”dra om ledningarna” i vår digitala infrastruktur handlar om tre saker:

- Modet att granska: Att våga ställa frågan ”vem saknas i den här datan?” innan systemet rullas ut.
- Makten att reglera: Att inse att teknisk innovation aldrig får ske på bekostnad av mänskliga rättigheter.
- Viljan att inkludera: Att förstå att mångfald i de team som bygger AI inte är en fråga om image – det är en fråga om kvalitet och säkerhet.

AI kommer att fortsätta spegla vår värld. Men när vi börjar se de blinda fläckarna kan vi också påverka vad den förstärker – eller inte förstärker. Frågan är inte hur smarta systemen blir. Utan vad vi väljer att bygga in i dem. Det är dags att börja koppla om.

