



KUNGL. FYSIOGRAFISKA SÄLLSKAPET I LUND

SYMPOSIET
**BEHÖVS VÅRA
HJÄRNOR?**

LUND 13 MAJ 2025

REDAKTÖR: MATS NYGREN

REDAKTIONSKOMMITTÉ:

PER ALM

MATS NYGREN

LUND 2026

Innehåll

Förord	3
Välkomna! <i>Ulf Söderlund</i>	4
AI – en gränsöverskridande historia <i>Carsten Peterson</i>	8
AI är mänsklighetens största ödesfråga <i>Olle Häggström</i>	xx
Hur tänker människor, djur och AI-system? <i>Peter Gärdenfors</i>	xx
AI som forskningens nya medarbetare <i>Fredrik Heintz</i>	xx
Har du köpt morötter? En fråga med variationer <i>Lena Andersson</i>	xx
AI – hot och möjligheter <i>Sophia Zackrisson</i>	xx
Generativ AI och högre utbildning <i>Thore Husfeldt</i>	xx
Läkekonst bland intelligenta algoritmer <i>Kristina Lång</i>	xx
Springer vi ifrån oss själva? <i>Peter Bovet Emanuel</i>	xx
Hur påverkar AI mänskligt arbete? <i>Julia Ravanis</i>	xx

Förord

AI, artificiell intelligens, är både ett hopp och ett framtidslofte – och ett hot mot vår framtid. Denna teknik har redan en roll i de krig som pågår just nu, där denna datateknik hjälper vapen att träffa sina mål i både Ukraina, Venezuela och Iran. Samtidigt har AI redan en viktig funktion som kartläggare av tumörer i våra kroppar. AI kan åt oss framställa bilder och filmer på händelser som aldrig har hänt i verkligheten. Vad händer då med vår verklighetsuppfattning och våra föreställningar?

Dessa frågor diskuterades på Rausingsymposiet i Lund den 13 maj 2025, ett heldagssymposium inför intresserade akademiker, som i vanlig ordning hölls på Grand Hotel. Här berättade forskare om grunderna för AI-tekniken och hur den utvecklats. En talare varnade för att AI kan bli mänsklighetens bokstavliga undergång, om vi inte ser till att skydda oss mot den. Andra värjde sig genom att påpeka att det är långt kvar innan AI kan agera utanför skärmarnas värld. En humanist visade att det mänskliga språket är så mycket rikare än maskinernas. Å andra sidan pekar pedagoger på att AI redan är överlägsen oss på att hantera just språk. Jag har redan nämnt den militära utvecklingen, som också diskuterades på symposiet. Där det kan vara svårt styra en AI som är snabbare än mänskliga soldater – och som står i ständig beredskap.

På samma sätt som språklärare kanske blir överflödiga i framtiden, kan läkare, sekreterare, tillverkningsarbetare och många andra yrkesgrupper ersättas av robotar med AI-hjärnor. Vem försörjer i så fall oss människor – och vilken roll får vi i det samhället?

Mats Nygren

Välkomna!



iStock™
Credit: mtcurado



Eric Warrant

Preses Kungl Fysiografiska
sällskapet i Lund 2025

Ärade ledamöter, moderatorer, föredragshållare, gäster och åhörare! Välkomna till årets Rausing-symposium i regi av det Kungliga Fysiografiska Sällskapet i Lund. Symposiet är det 16:e i en serie av årliga symposier, som har möjliggjorts av en generös donation av Ruben Rausing's fond, för vilken vi är mycket tacksamma. Dessa symposier har en tradition av att behandla ett högaktuellt ämne, som är lika viktigt för akademien som för samhället i stort. Årets tema – artificiell intelligens, eller AI – är inget undantag.

AI:s kraft och påverkan har ökat explosionsartat under endast ett fåtal år. Numera finns det ingen del av den mänskliga existensen som inte berörs i någon form av AI. Det är en utveckling som ger oss enorma möjligheter men samtidigt också väsentliga risker och faror. Vilka konsekvenser har en alltmer kraftfull AI för mänsklighetens framtid? Hur kommer AI att påverka vår utbildning, vår forskning, vår hälsovård, vårt arbetsliv, vår media och vårt försvar? Kan artificiell intelligens någonsin överstiga

mänsklig intelligens, och om så skulle sker kommer det att påverka oss negativt?

Och nu när AI alltmer används för att sudda ut gränserna mellan sanningar och osanningar och mellan verklighet och fiktion, hur ska vi hantera det som samhälle? Frågorna kring AI och mänsklighetens framtid är många men i nuläget är svaren få. Under dagens symposium försöker vi komma närmare dessa svar. Våra skickliga föredragshållare kommer att hjälpa oss med detta.

Ännu en gång, varmt välkomna till årets Rausingssymposium!



AI – en gränsöverskridande historia



**Carsten Peterson,
moderator**

Professor i teoretisk fysik,
Lunds universitet

Innan vi går in på ämnet för dagens symposium, *AI – hot och möjligheter*, inleder vi med en mycket komprimerad beskrivning av ämnesområdets utveckling från dess begynnelse till idag. Av utrymmesskäl anges inga referenser – fokus ligger på flödet av idéer och deras implementeringar.

KOMPLETTERING AV HOMO SAPIENS

Människan är dålig på att hantera stora beräkningar och invecklad logik. Därför uppstod tidigt behovet av att utveckla hjälpmedel – och så småningom datorer. Det första verkliga försöket att konstruera en beräkningsmaskin gjordes av Charles Babbage år 1822. Hans maskin, en mekanisk anläggning med hålkort och driven av en ångmaskin, förverkligades aldrig, men anses i efterhand ha varit fullt fungerande om den byggts.

Det stora genombrottet kom under andra världskriget, då behovet av att utföra avancerade beräkningar inom termodynamik, ballistik och kärnfysik blev akut bland annat för utvecklingen av atombomben. År 1945 färdigställdes ENIAC, världens första elektroniska dator. Den var byggd med vakuumrör

som förstärkare och kunde utföra beräkningar i en hastighet som vida överträffade allt som tidigare varit möjligt. Även i Sverige tog utvecklingen fart. Här i Lund fick vi SMIL (Siffermaskinen i Lund) år 1956, också den vakuumrörsbaserad. Vid samma tid började dock transistorerna ersätta vakuumrören, vilket gjorde datorerna både mindre, snabbare och mer driftsäkra.

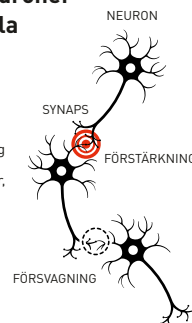
På 1960- och 70-talen följde nästa steg med integrerade kretsar och mikroprocessorer, grunden för de moderna datorer vi använder idag. Genom hela denna utveckling har den så kallade von Neumann-arkitekturen varit central med en dator organiserad i tre huvuddelar: den centrala beräkningsenheten (CPU), minnet, samt in- och utkommunikation (I/O). Programmeringen har förändrats i språk och verktyg, men principerna är desamma. Själv hade jag nöjet att skriva program i Algol för SMIL redan på 1960-talet. Dessa program skulle i princip fortfarande kunna köras på dagens datorer.

HÄRMANDE AV HOMO SAPIENS

Människan är däremot mycket bra på mönsterigenkänning, associationer och intuitiva slutsatser – förmågor som länge verkade svåra att överföra till datorer. Skulle det vara möjligt att få datorer att hantera även dessa problemställningar – och i vissa fall till och med överträffa Homo sapiens? Med denna målsättning etablerades år 1956 ämnet artificiell intelligens (AI). Initiativet togs framför allt av matematiker och logiker, vilket gjorde att det tidiga arbetet fokuserade på regelbaserade metoder. Målet var att formulera exakta regler och algoritmer, som kunde härleda fullt transparenta resultat. När det gällde mönsterigenkänning var framgångarna dock måttliga. Regelbaserad logik fungerade bra för schackproblem eller matematiska satser, men den räckte inte för att tolka bilder, språk eller komplexa situationer. Man skulle behöva gå utanför den strikta logikens gränser, mot lärandemetoder som kunde hantera osäkerhet och statistik.

Naturliga neuroner och artificiella

Neuronerna i hjärnan är levande celler som har ett avancerat inre maskineri. De kan skicka signaler till varandra genom synapser. Vid inlärning stärks kopplingarna mellan vissa neuroner, och andra försvagas.



Artificiella neuronnät är uppbyggda av noder som kodas med ett värde. Noderna har kopplingar till varandra. När nätverket tränas görs kopplingarna starkare mellan noder som är aktiva samtidigt, annars görs de svagare.

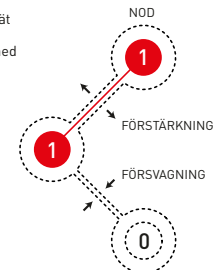


Fig. 1. Illustration: Johan Jarnestad/Kungl. Vetenskapsakademien

Kanske kunde AI-utvecklingen inspireras av hur hjärnan bearbetar information, inklusive hur minnet fungerar? Redan på 1940-talet hade en brokig skara forskare från olika discipliner, biologer, beteendevetare och matematiker, börjat studera hjärnan, inte för att bygga datorer utan för att förstå hur information bearbetas biologiskt. Inom neurovetenskapen var redan mycket känt. Hjärnans elementära byggstenar, neuronerna, är sammankopplade på ett komplext sätt där helheten är nyckeln till både sensorisk signalbearbetning och minnesfunktioner. Ett neuron fungerar i grunden som en förstärkare. Den tar emot signaler via synapser, summerar dessa, och skickar vidare en impuls om den sammanlagda signalen är tillräckligt stark. Kopplingarna mellan neuroner, axonernas synapser, kan förstärka eller hämma signaler, och det var dessa kopplingars styrkor, eller vikter, som man tidigt tolkade som centrala för minne och lärande.

ARTIFICIELLA NEURALA NÄTVERK

Efterhand förenklade man den biologiska komplexiteten i modellerna. Istället för att modellera de elektriska pulserna, beskrev man neuronerna som noder som kunde vara antingen *på* eller *av* (Fig. 1). Kopplingarna mellan dem representerades av vikter som antingen förstärkte eller försvagade inkommande

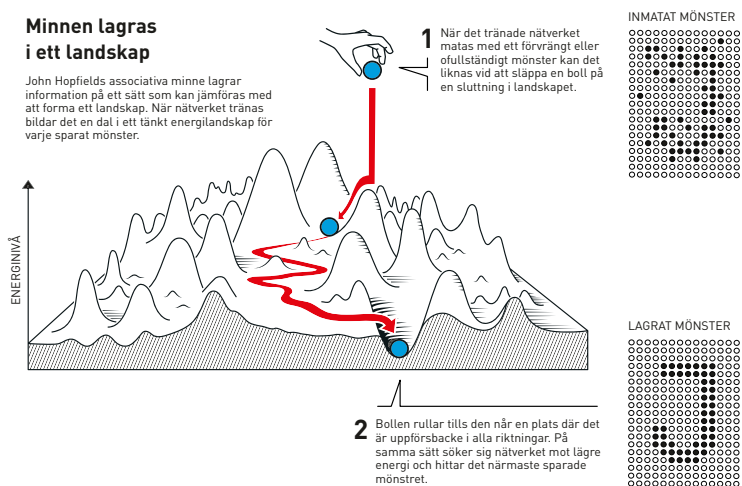


Fig. 2. Illustration: Johan Jarnestad/Kungl. Vetenskapsakademien.

signaler. Om summan av viktade signaler nådde över en viss tröskel aktiverades noden. Denna förenklade modell av biologisk signalhantering lånar sig väl till datorsimuleringar, och under 1970-talet växte ett stort intresse fram för att bygga artificiella neuronnätverk (ANN). Målet var att skapa datorer som kunde lära sig genom exempel snarare än att enbart följa explicita regler. Samtidigt fortsatte den biologiska forskningen med oförminskat fokus på hjärnans funktion, vilket gjorde att de två spåren, biologisk förståelse och teknisk tillämpning, utvecklades sida vid sida.

Samtidigt, under 1970-talet, kom en kraftig och inflytelserik kritik mot ANN från det av matematiker dominerade AI-samhället. Kritikerna menade att dåvarande ANN inte klarade av att lära sig vissa intrikata logiska samband. Detta var i sak korrekt, men problemet var mer tekniskt än fundamentalt och kom kort därefter att lösas. Trots detta fick kritiken stort genomslag och resulterade i en period av minskat förtroende och bristande finansiering för forskningen på området. Denna period brukar kallas för AI-vintern.

Den grundläggande skillnaden mellan regelbaserade system och neuronnät var dock tydlig. Regelbaserade system byggde på att man i förväg formulerade logiska regler som datorn följde. ANN å andra sidan lärde sig i stället strukturer och samband direkt från exempel från data. Ur ett vetenskapsteoretiskt perspektiv är detta inte särskilt märkvärdigt. Det kan jämföras med att anpassa en rät linje, det vill säga bestämma parametrarna, till en uppsättning datapunkter (inlärning), och därefter använda den linjen för att generalisera till nya punkter via interpolering. Samma princip gäller för ANN, som är mer komplexa med många fler dimensioner och icke-linjära samband, där vikterna utgör systemets parametrar.

TVÅ GENOMBROT

Trots det svala mottagandet brakade det loss för ANN under 80-talet. Man fick genombrott när det gällde två slags neuronstrukturer.

(i) *Återkopplade nätverk*, där signaler färdas både framåt och bakåt mellan neuronerna. Fysiker spelade en central roll i utvecklingen här genom att metaforiskt introducera en energifunktion med neuroners värde som variabler och vikterna dem emellan som konstanta faktorer. De senare relaterade till inlärda mönster eller minnen motsvarande funktionens minima. Denna energi är en monotont avtagande funktion av tiden, vilket innebär att lokala energiminima utgör stabila tillstånd. Om nätverket startas från ett initialt tillstånd motsvarande ett förvrängt mönster, så kommer dynamiken att driva systemet mot lägre energi, till en närbelägen dal. På detta vis kan ett förvrängt mönster korrigeras; ett korrekt mönster kan återskapas (Fig. 2). Detta kan liknas vid ett landskap där en "kula", exempelvis en textsträng, rullar ner till närmaste dalgång (minimum). Resultatet blev ett associativt minne – Hopfield-modellen. Ett felstavat ord kunde automatiskt attraheras till sin lagrade, korrekta motsvarighet, där inlärningen byggde på korrelationer

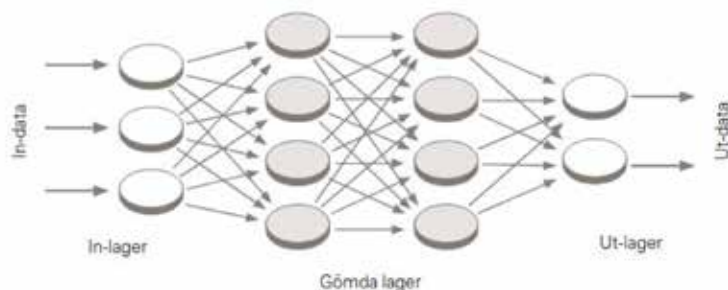


Fig 3. Enkelriktat nätverk med två lager av gömda noder mellan in- och ut-lagren. Illustration: C. Kjellstrand. Baserad på illustration av Johan Jarnestad/Kungl. Vetenskapsakademien.

mellan neuronerna. Hopfield-modellen utgjorde sedan grunden för Boltzmann-maskinen, där bland annat en "temperatur" (eller brus) infördes, vilket på ett naturligt sätt gav upphov till fördelningar kring lösningarna.

(ii) *Enkelriktade nätverk* (Fig. 3) fick under samma period sitt stora genombrott. Här flödar informationen i en enda riktning – från indata till utdata via ett antal "gömda" mellanlager. En typisk arkitektur kunde ta en bild som insignal och ge en binär utsignal: till exempel 1 om bilden föreställde en hund och 0 annars. Systemet lärde sig denna avbildning genom att vikterna mellan neuronerna successivt anpassades, när det visades många exempelbilder med kända svar. För att effektivt bestämma dessa vikter utvecklades metoden *backpropagation*, som bygger på kedjeregeln tillämpad på en felfunktion. Till skillnad från de återkopplade näten var de enkelriktade betydligt mindre beräkningskrävande. Detta bidrog till deras snabba genomslag, särskilt inom praktiska tillämpningar som bildklassificering.

EN "VINTER" TILL

Under 1990-talet blev det återigen svårt att finansiera forskning kring ANN. En bidragande orsak var den ökade konkurrensen

från statistiska metoder, som ofta uppfattades som mer pålitliga och matematiskt välgrundade. Samtidigt stötte ANN på problem med att hantera stora nätverk och relaterad överinlärning som begränsade deras generaliseringsförmåga. Dessutom saknades övertygande storskaliga tillämpningar. Förutom enklare uppgifter inom bildbehandling och vissa medicinska problemställningar fanns få exempel som kunde visa den breda nyttan av ANN. Detta bidrog till att forskningsintresset och forskningsresurserna minskade under perioden.

YTTERLIGARE TVÅ GENOMBROT

Två viktiga genombrott kom att förändra utvecklingen av artificiella neuronnät:

(i) *Den första utmaningen* gällde att begränsa det snabbt växande antalet parametrar, det vill säga vikterna i nätverken. Här spelade de så kallade falttningsnätverken (convolutional networks), introducerade 1989, en avgörande roll. Denna arkitektur reducerade antalet kopplingar mellan intelligande lager genom att införa lokala receptiva fält, inspirerat av neurovetenskapliga fynd. Utöver den praktiska fördelen med färre parametrar erbjöd metoden också en mer intuitiv tolkning av de gömda lagren. Noder i dessa kunde representera olika egenskaper i den inkommande signalen såsom kanter och andra visuella attribut inom bildbehandling. Därmed kunde nätverken hantera variationer i objektens position i en bild, vilket tidigare varit en stor utmaning.

(ii) *Det andra genombrottet* gällde hur djupa nätverk med många gömda lager skulle kunna tränas. Fullt konnektiva arkitekturer, där varje neuron i ett lager är kopplad till alla neuroner i nästa lager, hade länge visat sig svåra att optimera med vanliga metoder som backpropagation. Under 2000-talet introducerades därför tekniker för att förträna nätverken lager för lager. En viktig innovation var att använda varianter av Boltzmannmaskinen, där korrelationer mellan gömda noder kunde etableras innan

nätverket finjusterades med backpropagation. Detta gjorde det möjligt att träna betydligt djupare modeller än tidigare. Senare kunde den beräkningsintensiva Boltzmann-maskinen ersättas med enklare arkitekturer, vilket ytterligare ökade metodens användbarhet.

Dessa framsteg markerade början på ett nytt paradigm: *deep learning*. Med djupa nätverk flyttades fokus från enskilda noders värden till att modellera fördelningar av dessa. Detta öppnade i sin tur för utvecklingen av så kallad generativ AI, där nätverken inte enbart analyserar data utan också kan skapa nya, syntetiska data – exempelvis bilder, ljud och text.

Under 2010-talet börjar Google och andra IT-giganter visa stort intresse. Dessa har ju redan en intern forsknings- och utvecklingskultur på plats samt enorma investeringsresurser. Toppforskare anställs här på hel- och deltid. Idéer från dessa kan då snabbt testas och utvärderas med hundratals programmerare – resurser som universiteten saknade. Dessutom har dessa företag tillgång till stor beräkningskapacitet samt, inte minst, stora mängder träningsdata, där mycket ligger i molnet. Implementeringen av *deep learning* ledde till flera spektakulära och trovärdiga tillämpningar. Under denna tid slogs regelbaserad AI ihop med ANN med samlingsnamnet AI – en produktiv konsolidering.

SPRÅKMODELLER

Språket är en framträdande ”egenskap” hos *Homo sapiens*. Längre hade regelbaserade system inom AI varit engagerade i att utveckla modeller för språk, ofta med lingvistik som viktig ingrediens. Dessa modeller, som byggde på explicita grammatiska regler och semantiska strukturer, visade sig svåra att skala upp till den komplexitet som naturligt språk innebär. Inom ANN hade man sedan början av 2000-talet också varit aktiv här, då med att från ett givet ord eller sekvens av ord förutsäga nästa ord i en mening – återigen ett lärande system.

Ett avgörande genombrott kom dock med introduktionen av transformer-arkitekturen 2014 och 2017. Denna förändrade fundamentalt hur språkmodeller tränas och används. Till skillnad från tidigare arkitekturer möjliggör *transformern* effektiv hantering av långa sammanhang och parallell bearbetning, vilket dramatiskt förbättrade både prestanda och skalbarhet. Transformern har därefter generaliserats långt bortom språkteknologi. Samma principer används idag för bildbehandling, analys av tabulära data och många andra typer av uppgifter. Arkitekturen har blivit en central komponent i nästan alla moderna AI-tillämpningar och ligger till grund för stora språkmodeller såsom ChatGPT, där mänsklig interaktion med AI har nått en helt ny nivå.

ÅTERKOPPLING TILL BIOLOGIN

De flesta framgångar inom AI hittills har datormässigt uppnåtts genom digitala komponenter. Implementationerna har i regel följt en von Neumann-struktur, ofta med specialiserade arkitekturer för parallell bearbetning som grafikprocessorer (GPU:er). De senare, inom vilka företaget Nvidia länge haft en ledande position, har gett stora prestandavinster.

ANN är ju i grunden inspirerad av biologin. I hjärnan utgör kretsarna av neuroner analoga system, medan dagens ANN körs i digital form. Att bygga verkligt analoga kretsar har gått långsamt. En viktig orsak är att det varit mer kostnadseffektivt att utgå från redan etablerade digitala plattformar med breda användningsområden, snarare än att utveckla helt nya teknologier. Med analoga kretsar skulle fysikens ekvationer direkt kunna utnyttjas för att lösa problem, på samma sätt som hjärnan gör – utan mellansteget av programmering och kompilering. En analog förstärkare har dessutom inte samma höga precision som en 64-bitars digital komponent, vilket gör att energiförbrukningen kan bli betydligt lägre. Detta är särskilt relevant för inbyggda system, exempelvis medicinska sensorer inuti kroppen.



Frank Wollheim

Utöver energieffektivitet finns även stor potential i att använda analoga kretsar för att simulera biologiska neurala system. Analoga kretsar erbjuder både tekniska fördelar och verktyg för biologiska simuleringar – men de står ännu i skuggan av den digitala dominansen.

SINNRIKA ALGORITMER

Utvecklingen av AI med ANN som grundpelare initierades av beteendevetare och neuroforskare, varför lärande system tidigt var i fokus snarare än logisk programmering. Algoritmer utvecklades tidigt för inlärning som successivt övervann kritik från matematiker och sedermera dataloger. Under lång tid var det dock en stor utmaning att träna stora nätverk, dels på grund av begränsad datorkraft och dels för bristen på stora datamängder. Dessa hinder övervanns till stor del genom förbättrade och sinnrika algoritmer inom så kallad *deep learning*, samt genom att de stora IT-företagen engagerade sig i området. När det mer formella AI-samhället anammat ANN, kom flera kraftfulla förbättringar, t.ex. inom språkmodeller. Mekanismerna bakom de senare styr mycket av utvecklingen idag.

Tillämpningarna går nu framåt med stormsteg och enastående resultat alstras till mänsklighetens gagn inom ett vitt spektrum av ämnesområden. Dessa framgångar måste vägas mot faran att *Homo sapiens* tappar kontrollen över sitt öde. Detta är frågeställningen för dagens symposium.



AI är mänsklighetens största ödesfråga

Gudinnan Fortuna



Olle Häggström

Professor i matematisk statistik vid Chalmers tekniska högskola

Jag har stor respekt för att det finns andra frågor än AI – exempelvis klimatförändringarna – som kan göra anspråk på att vara stora ödesfrågor för mänskligheten, men de senaste åren har jag kommit att landa i att frågan om hur vi hanterar AI-utvecklingen är den största och mest akuta av dem alla. Jag skall förklara varför, men vill börja i en mer jordnära ände.

Kan AI bli bättre än läkare på medicinsk diagnosticering? Mycket har sagts och skrivits om detta ämne, och vi skall höra en hel del om saken i Kristina Långs föredrag i eftermiddag, men här vill jag som exempel framhålla två aktuella studier i tidskriften *Nature* som pekar mot att svaret kan vara ja. De är begränsade till en ganska snäv kontext med enbart diagnosticering via textinterface, men det är ändå intressant att det AI-system som studeras presterar bättre än mänskliga läkare ställda inför samma uppgift. Studiernas begränsningar gör att det är på tok för tidigt att hävda att läkarna redan idag skulle vara akterseglade av AI, men utvecklingen går snabbt och vi kan komma att nå den dag då klockan klämtar för läkarna snabbare än vi annat.

gränssnitt?



Den stackars
läkaren-i-loopen.

I frågor om AI och automatisering av beslutsfattande talar man ofta om det så kallade *human in the loop-idealet*. Detta går ut på att AI inte skall *ersätta* utan blott *assistera* mänskligt beslutsfattande, och att människan alltid skall ha sista ordet. Extra näring för detta ideal fås ur idén om att kombinationen människa plus AI i många fall presterat bättre än AI ensam – även i fall där AI överglänser människan ensam.

EXEMPLET SCHACK

Det paradexempel på detta fenomen som oftast anförts av AI-entusiaster kommer från schacket. Kraftmätningen mellan människa och maskin på schackets område har ansetts avgjord ända sedan 1997, då dåvarande schackvärldsmästaren Garri Kasparov förlorade en uppmärksam match mot IBM:s schackprogram Deep Blue. Redan året efter tog han initiativ till den nya spelformen kentaurschack, där mänskliga spelare fick lov att konsultera ett schackprogram innan de utförde sina drag. Så länge teamet människa plus maskin var bättre än maskinen ensam förblev spelformen intressant, men intresset kollapsade i

Kasparov under
den symboliskt
ödesmättade
schackmatchen
mot Deep Blue
1997.



slutet av 2010-talet då det blev snudd på omöjligt att överträffa en motståndare som valde att slaviskt följa datorns rekommendationer, och det mänskliga bidraget till spelformen därför blivit irrelevant.

Ett liknande fenomen påvisas i den ena av de nämnda *Nature*-studierna. Läkare med tillgång till forskarnas AI-system uppvisade *sämre* precision i sina diagnoser än AI-systemet ensamt. Under sådana omständigheter, där det mänskliga deltagandet aktivt försämrar utfallet, är det svårare att försvara människan i loopen. Vi är som sagt inte där än, men om och när det inträffar gissar jag att vi ändå in i det längsta kommer att insistera på att viktiga vårdbeslut skall fattas av mänskliga läkare. Deras situation blir då föga avundsvärd, då de sitter med det formella ansvaret för ett beslut som de inte vill gripa in i, eftersom de inser att det vore kontraproduktivt, och som de till följd av AI-systemens black box-egenskap kanske inte ens begriper grunderna för.

VI TAPPAR MAKT

Motsvarande utveckling kan vara på väg att ske i snart sagt alla samhällssektorer. Den dynamik som därmed uppstår har

studerats av forskare under benämningen *gradual disempowerment* (gradvis förlust av makt). Effektivitetskrav leder allt fler aktörer att plocka bort människan ur loopen, och de företag som av exempelvis idealistiska skäl undviker det slås ut i konkurrensen. Den logiska slutpunkten för en sådan utveckling är att människan inte längre deltar i styrningen i samhället, utan den saken helt tagits över av AI, och i det läget kan vi inte längre påverka framtiden. Vi behöver därför snabbt få till stånd en välinformerad diskussion om vart vi är på väg och vilket slags samhälle med AI vi bör eftersträva – innan det är för sent.

En typ av arbetsuppgift där AI gjort särskilt dramatiska framsteg de senaste åren och som kan komma att få stor betydelse för den fortsatta utvecklingen som helhet är kodning och mjukvaruutveckling. AI-systemens förmåga att skriva korrekt kod och i övrigt lösa uppgifter riktigt är starkt avhängig uppgiftens omfattning. I en rapport från AI-säkerhetsorganisationen METR i mars i år studeras hur denna förmåga utvecklats över tid. Det visar sig att omfattningen – mätt i tidsåtgång för en mänsklig expert – som AI klarar av har ökat från enstaka sekunder 2019 till cirka en timme idag. Ökningen är exponentiell, med en observerad genomsnittlig fördubblingstid på sju månader, och om man extrapolerar den trenden blott ett par-tre år in i framtiden blir resultatet dramatiskt. Sådan kurvanpassning inbegriper givetvis stora osäkerheter, men ser man till hur modellerna förbättrats från 2024 och framåt verkar det snarast som att utvecklingen är på väg att gå ännu fortare.

SUPERINTELLIGENS 2027?

Det är bland annat den sortens data som ligger till grund för den gedigna rapporten *AI 2027*, utkommen i april i år och författad av en kvintett forskare med den avhoppade OpenAI-medarbetaren *Daniel Kokotajlo* i spetsen. Rapporten är det mest ambitiösa och bästa som hittills skrivits vad gäller detaljerade förutsägelser av kommande års AI-utveckling. Osäkerheterna

är som sagt stora, men successivt och månad för månad arbetar de fram vad de ser som det mest sannolika förloppet. Centralt i detta förlopp är hur AI, till följd av den utveckling som bland annat METR-rapporten påvisat, år 2027 når en punkt där den är en lika skicklig AI-utvecklare som dagens främsta sådana av kött och blod. Tack vare att de ledande AI-företagen då kan sätta hundratusentals eller miljontals sådana AI i arbete leder detta på några få månader till så kallad superintelligens – AI som vida överträffar människan över hela spektret av relevanta förmågor.

Vad som därefter kan förväntas hända är långt ifrån självklart, och i själva verket har förloppet i Kokotajlos rapport en bifurkation i slutet av 2027, där det delar sig i två, varav det ena landar i utplåning av mänskligheten redan 2030, och det andra går betydligt bättre. Mänsklighetens undergång kan låta överdrivet dramatiskt, men vad superintelligens i praktiken innebär är ju att vi har skapat en ny art som petar ned oss till en andraplats bland planetens mest kapabla och intelligenta arter, och vi bör inte ta för givet att en sådan situation automatiskt slutar lyckligt för den art som akterseglats.

Här kan det vara värt att gå tillbaka och höra vad några tidigare tänkare på området sagt. Datavetenskapens fader *Alan Turing* gick bort under förfärliga omständigheter redan 1954, ännu inte 42 år fyllda och två år innan termen artificiell intelligens skulle uppstå som namnet på ett forskningsområde. Merparten av hans livsgärning var av matematisk och teknisk natur, men de sista åren av sitt liv tillät han sig enstaka mer filosofiska utblickar och reflektioner kring framtiden för den teknologiutveckling han så starkt bidragit till att sätta igång. I en ofta citerad passage från 1951 spekulerar han över vad som kan hända om vi lyckas skapa maskiner som kan tänka på ett sätt som liknar människors. Kokotajlo och andras mer sentida idéer på samma tema kan delvis spåras tillbaka till hur Turing här ser framför sig att maskinerna vid en viss nivå kan börja förbättra sig själva utan behov av vidare mänsklig inblandning,



Thore Husfeldt
och Olle Hägg-
ström i panel-
samtal med
Christiane Vejlø
i Köpenhamn
2019.

varvid utvecklingen kan väntas accelerera kraftigt, så att ”vi så småningom bör vänta oss att maskinerna tar över kontrollen”.

INGEN LYSSNADE

Vilket inflytande hade då dessa ödesmättade ord på den AI-forskning som snart skulle komma igång? Att döma av det ambitiösa och oförvägna arbete som AI-forskarna kom att bedriva, vilket under det halvsekel som följde efter att Turing formulerade sina rader var nästan helt befriat från risktänkande, förblev hans påverkan på denna punkt länge nära nollnivån. Ett av mycket få undantag under denna tidsperiod var *Norbert Wiener*s remarkabelt framsynta text *Some moral and technical consequences of automation* från 1960, där författaren resonerar kring en tänkt framtida situation, där vi skapar en maskin så kraftfull att vi när vi väl satt igång den saknar förmåga att stänga av den. Han betonar vikten av att säkerställa att ”det syfte vi byggt in i maskinen är det syfte vi eftersträvar snarare än någon färgstark imitation”. Dessa ord skulle faktiskt kunna fungera väl som programförklaring för det område som inte uppstod

förrän först många årtionden senare, men som idag benämns *AI alignment*, vars avsikt är att se till att de första riktigt kraftfulla AI-systemen har mål och värderingar som i tillräcklig mån prioriterar mänsklig välfärd och blomstring och mer allmänt är i linje med mänskliga värderingar.

AI alignment-pionjären *Eliezer Yudkowsky* kan med blott en mild överdrift sägas egenhändigt ha lagt grunden för området under 00-talet. I en inflytelserik artikel från 2008 beskriver han det han bedömer vara default-scenariot ifall vi misslyckas med eller helt enkelt ignorerar AI alignment: ”AI:n hatar dig inte, ej heller älskar den dig, men du består av atomer som den kan ha annan användning för”.

En sådan situation vill vi givetvis inte hamna i, och därför behöver vi lösa AI alignment i tid. Hur lång tid har vi då på oss? Ingen vet säkert, och det enda omdömesgilla är att medge att stor osäkerhet föreligger, men jag menar att vi bör ta på allvar den i AK-ekosystemet i San Francisco och Silicon Valley alltmer utbredda uppfattningen att superintelligent AI kan bli en realitet inom en tidsrymd som mäts i enstaka år snarare än decennier.

MÄNSKLIG UTROTNING?

Kokotajlo-rapporten är ett uttryck för denna uppfattning, och den tidigare styrelseledamoten i Open AI, *Helen Toner*, vittnade i ett senatsförhör i Washington den 17 september 2024 om hur vanlig den är bland insiders på de ledande AI-företagen. Hon tillade att ”många av dessa bedömare tror att om de lyckas att bygga datorer som är lika smarta som människor eller eventuellt långt smartare, så kommer tekniken att som minst bli extremt omstörtande, och som mest leda till bokstavlig utrotning av mänskligheten”.

Det finns såklart andra som ställer sig avvisande till sådana bedömningar, ibland med hänvisning till att dagens AI inte besitter något som överhuvudtaget förtjänar att kallas intelligens, och jag vet att flera av er här i lokalen har denna uppfattning. Argumenten varierar, men jag har ännu inte stött på något som håller för

närmare granskning. Två särskilt vanliga argument i samband med den kanske viktigaste AI-tekniken i dagens frontlinje – deep learning-drivna stora språkmodeller – är för det första ”allt de gör är att prediktera nästa ord i internettext”, och för det andra ”allt som händer innerst inne i de neurala nätverken är den triviala matematiska operationen matrismultiplikation”. Båda två är exempel på det argumentationsfel som kan benämnas *reductio ad reductem*, och som innebär att man förnekar möjligheten till emergens av komplexa fenomen ur sammansättningen av enkla, eller med andra ord att man ur observationen att helheten kan sönderdelas i enkla beståndsdelar drar slutsatsen att helheten inte finns.

Varför hävdar jag att det är ett argumentationsfel? Det enklaste sättet att inse det är att tillämpa argumentet på någon erkänt intelligent person som exempelvis Albert Einstein, och konstatera att Einsteins hjärna var uppbyggd av atomer och elementarpartiklar som var och en är stendum, varför sagda hjärna omöjligt kan ha varit ett säte för intelligens. Eftersom vi ju var överens om att Einstein faktiskt var intelligent kan vi därmed konstatera att *reductio ad reductem* inte är någon giltig slutledningsprincip, varvid argumenten om att ordprediktion och matrismultiplikation inte räcker till som beståndsdelar i intelligenta system kollapsar. I en artikel publicerad 2023 går jag igenom en rad andra försök att visa att stora språkmodeller inte besitter resonemangsförmåga eller intelligens. Jag destruerar dem alla med samma grepp, nämligen att påvisa att om argumenten vore riktiga skulle de även medföra icke-existensen av mänsklig intelligens – en slutsats vi vet är falsk, varför vi kan konstatera att argumentet inte håller.

Den kanske allra vanligaste typen av argument för de stora språkmodellernas icke-intelligens är att peka på fall där de svarar korkat på någon fråga eller på annat vis misslyckas med någon uppgift som borde vara enkel, och hävda att ingen verkligt intelligent entitet skulle göra bort sig så kapitalt. Därmed, påstås

det, är vi uppenbarligen långt från det som brukar kallas *AGI – artificiell generell intelligens* – vilket brukar definieras som en AI som presterar på mänsklig nivå eller högre inom hela spektret av relevanta kognitiva förmågor. Men med det argumentet har vi lagt ribban för högt – så högt, vågar jag påstå, att ingen här i lokalen kan göra anspråk på intelligens, för upp med en hand alla som aldrig någonsin sagt något dumt eller misslyckats med något enkelt!

VILKEN KOMPETENS?

Det fanns en tid när AGI-begreppet var ett viktigt pedagogiskt verktyg, men jag tror att begreppet har överanvänts på senare år (något jag själv erkänner mig medskyldig till) och att vi idag nått en punkt där det bidrar mindre till klarhet än till förvirring, genom dess betoning på att AI skall klara *allt* som människan kan. Det uppmuntrar till det felslut jag här diskuterat, om att AI kan avfärdas som meningslös så snart vi ser att den misslyckas med någon enskild enkel uppgift. Dessutom leder för de flesta praktiska frågeställningar en överbetoning av AGI till fel fokus. Tag till exempel den fråga jag här intresserar mig för mest av allt: den om huruvida och i så fall när AI kan bli kapabel att utmana människan om herravälde över samhället och vår planet. Det viktiga för denna frågeställning är såklart inte när AI överträffar oss i *allt*, utan när den blir tillräckligt mycket bättre på ett tillräckligt brett spektrum av kompetenser som är relevanta för maktövertagande. Den språkliga kompetens som mycket av dagens AI-utveckling kretsar kring, tror jag kan vara viktig här med tanke på vilken stor andel av all maktutövning som i dagens samhälle äger rum via språkhandlingar. Måhända ännu viktigare kan visa sig bli den mjukvaruutvecklingskompetens jag varit inne på ovan. Däremot tror jag inte alls att det är nödvändigt för AI att kunna knyta ett skosnöre innan den är redo att ta över världen.

Men, som antytts, tror jag att det totala AI-maktövertagandet är något vi bör undvika, och vi bör åtminstone se till att vi har



Vicepresident JD Vance under sitt vettlösa tal i Paris vid AI-toppmötet i februari 2025.

löst AI alignment innan vi skapar superintelligens. Just nu ser det besvärligt ut med tanke på den rasande utvecklingskurva AI befinner sig på, samtidigt som problemet med AI alignment gäckar oss och ingen har någon övertygande plan för hur det skall lösas.

NÖDBROMS

För att ge AI alignment-forskningen en chans att hinna ikapp tror jag att vi behöver dra i nödbromsen för utvecklingen av de allra mest kraftfulla AI-systemen. Detta försvåras dock av den kapplöpningssituation som föreligger, både mellan enskilda AI-företag och mellan länder (främst USA och Kina). Det allmänt hårdnande internationella klimatet under Trumps andra presidenttillträde gör inte heller saken lättare. Ett lågvattenmärke för den globala AI-diskursen nåddes vid toppmötet AI Action Summit i Paris i februari i år, där säkerhetsfrågor sopades under mattan samtidigt som toppolitiker bjöd över varandra i vilka mångmiljardbelopp de avsåg satsa på AI-utveckling. Värst av allt

var hur den amerikanske vicepresidenten JD Vance i sitt anförande uttryckte oförblommerat förakt för AI-säkerhet, då han slog fast att han "inte var där för att tala om AI-säkerhet" och att "vår AI-framtid inte erövrar genom att oja sig över säkerhet utan genom att bygga". Hans förhoppning lite längre fram i samma tal om att "AI-ekonomin kommer att [...] transformera den värld som består av atomer" ger, för den som minns Yudkowskys ovan citerade "one-liner" om AI och atomer, en isande rysning längs ryggraden.

Detta säger något om vilka krafter vi behöver övervinna, om vi skall få ordning på AI-utvecklingen och styra den i för mänskligheten mer gynnsam riktning jämfört med vart vi idag verkar vara på väg. Men framtiden är inte ristad i sten, och jag tror fortfarande att det är möjligt att förhindra en AI-katastrof. Något som skulle förbättra oddsen ytterligare vore, om vi lyckas mobilisera den folkopinion mot skapandet av övermänskligt intelligent AI, som enligt diverse opinionsundersökningar verkar föreligga. Så hjälp gärna till att sprida budskapet!

Bilderna är framställda av ChatGPT på beställning av Olle Häggström.

Joakim Wernberg, socioekonomiska teknikstudier, LTH: Tre korta funderingar: 1) Vi talar om AGI som att kunna utföra alla slags uppgifter. Men att utföra en uppgift är inte samma sak som att ha en förmåga. Om AI och du spelar schack så gör ni egentligen två olika typer av arbete fysikaliskt. Så är det ett race? Blir AI hela tiden bättre än oss eller kan den göra samma saker som oss och kanske bättre? 2) Och om AI blir bättre än oss på alla förmågor, menar du att den får agens på en och samma gång? För om vi tittar på biologin och evolutionen så ser vi först en intention hos biologiska varelser, därefter rationalitet och därefter sociala normer. Det verkar vara en förhärskande modell. Var skulle du säga att AI befinner sig i den utvecklingen? Och till sist, 3): Du spelar ju schack. Grattis till bronsmedaljen här om dagen! Anledningen att kentaurmatcherna las ner är inte att begreppet inte funkar. Man kan inte spela schack idag helt utan datorer, den tävlingsformen blev inte spännande längre. Men vi spelar ju fortfarande schack – och det är väl bra? Efterfrågesidan i ekonomin, det är vi som bestämmer över den. Den avgör ju hur vi använder tekniken.

Olle Häggström: Ja, jag blev trea i veteran-SM i snabbschack! Låt mig börja svara i fel ände med schacket. Jag tycker att vi i schacket har hanterat AI-framstegen väldigt bra och konstruktivt. Jag är i första hand schack-fan och lägger ohyggligt många timmar på att följa elitens matcher. Jag vill se den mänskliga aspekten, som när Magnus Carlsen svettas över ett offer, inte datorer som spelar mot varandra, trots att de spelar bättre schack.

Och det är den efterfrågan på det rent mänskliga som kommer att rädda oss på arbetsmarknaden, åtminstone

under en övergångsperiod (när AI kan utföra allt fler arbetsuppgifter). När det gäller din fråga om förmågor vill jag se det rent instrumentellt. Maskinerna tänker lite annorlunda än oss men löser sina uppgifter. När det gäller agens har de väl varit relativt svaga, men många talar om att 2025 kommer att bli agensens genombrottsår. Redan förra året släpptes AI-modeller som vi kan låta ta kontrollen över våra laptops, sköta våra flygbokningar etc. Det sker i dagsläget mycket taffligt och otillförlitligt, men jag tror här som på så många andra områden på en snabb utveckling. Agensen är nog snabbt på väg. Det är prioriterat från Open AI:s och Anthropic:s sida. Då har jag åtminstone berört dina tre frågor.

Ulf Kristoffersson, klinisk genetik: Omkring 1975 tog ett antal forskare initiativ till en konferens i Asilomar om nya biologiområden och risker med den forskningen. Man la ett tvåårigt moratorium medan man diskuterade säkerhetsregler. Jag vet att motsvarande har kommit upp när det gäller AI. Hur har det tagits emot av det vetenskapliga samhället idag? Har man varit följsam som då, eller har det varit ett intressedokument?

Olle Häggström: Det är ett intressedokument med vackra ord och följsamheten bland AI-utvecklarna är väldigt nära noll. Vi behöver hitta mer bindande regleringar för att hantera detta, det är min övertygelse. Det är bra att Asilomar-dokumentet finns men det räcker inte.

Bengt Sjölund, neurofysiolog, Lund: Jag upplevde den här 70-talsvågen redan som ung doktorand. När man nu använder AI-system baserade på språkmodeller och vi ber den skriva till exempel en review, så visar det sig att en del av nyckelreferenserna i texten har systemet själv hittat på, de finns inte! Och vi kan själva hitta på, fantisera, om en massa

saker. Vad händer när AI-systemen börjar fantisera om det som inte finns?

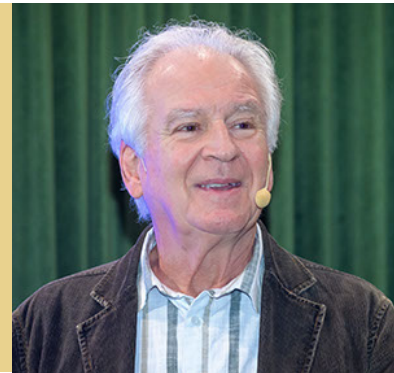
Olle Häggström: Ja, de gör ju det. Det är det så kallade hallucinationsfenomenet. Det är ju en annan högt prioriterad fråga för de ledande AI-utvecklarna. Vi är ju mindre benägna att använda deras AI-modeller om vi inte kan lita på dem. Så de jobbar hårt på att minska dessa problem och de lyckas i viss mån, men det verkar väldigt svårt att utplåna detta fenomen. Kanske är situationen inte så annorlunda än om du arbetar med kolleger, doktorander etc. Kan du lita på vad de säger? Vi kan ju också fabulera.

Eric Warrant, zoolog, Lund: Om AI vill utrota oss, hur ska det gå till?

Olle Häggström: Låt mig ge ännu en schack-analogi: Om jag spelar mot Magnus Carlsen, tidigare schackvärldsmästare, så vet jag med mycket stor säkerhet att han kommer att vinna. Men jag vet inte hur. Om jag kunde förutse hans drag vore jag ju en minst lika bra schackspelare som han är. Så utmaningen att jag skulle förutse vad en superintelligent AI ska göra är lite orättvis. Men i alla fall, jag tror inte på humanoida robotar a la Terminator-filmer som skjuter på folk. Robotiken ligger lite efter andra områden än så länge. Men i stället för robotar kan AI använda oss människor till att gå dess ärenden i den fysiska världen. Detta kräver att den behärskar övertalning och social manipulation, förmågor som vi redan ser starka tecken på hos dagens AI.



Hur tänker människor, djur och AI-system?



Peter Gärdenfors

Professor em, kognitions-
forskning, Lund

De senaste åren har utvecklingen inom artificiell intelligens (AI) gått oerhört snabbt. Nya system med allt mer överraskande förmågor presenteras i hög fart. En del forskare varnar för att systemen snart kommer att utveckla superintelligens och på något sätt ta över världen. Flera forskare tror emellertid att vi snart inte bara har AI-system som uppvisar intelligens inom ett snävt område, utan att det kommer att uppstå system med artificiell generell intelligens (AGI). Med detta menas system som blir skickligare än människor inom huvuddelen av de områden som vi behärskar.

Jag ska argumentera för att det är långt kvar tills vi når AGI. Jag vill försöka bringa reda i begreppen. Mitt mål är att visa att människors och djurs tänkande innehåller ett stort antal komponenter, medan AI-system är specialiserade till ganska snäva områden.

KAN AI TÄNKA?

Låt mig börja med frågan: Kan AI-system tänka? Jämför frågan med: Kan flygplan flyga? Det verkar självklart att flygplan flyger. I många avseenden överträffar de fåglar, fladdermöss och andra flygande djur: Flygplan kan flyga längre och snabbare, de kan bära stora mängder last och de behöver inte vila. Men flygplan fungerar enligt helt andra principer än fåglar. Fåglar flyger därför att det hjälper dem i deras strävan efter att överleva och fortplanta sig. Flygplan strävar inte efter något. De är verktyg för människans mål. Fåglars flygande är också mer flexibelt än flygplanens. På ett analogt sätt kan man säga att AI-systemens ”tänkande” i många avseenden överträffar människors och andra djurs. Systemen är snabbare, har bättre minne och kan lösa många problem som människor inte kan. Tänk bara på schackprogrammen, som utan större problem vinner över alla mänskliga schackspelare! Men AI-systemens tänkande fungerar på helt andra sätt än människornas. Människors tänkande har utvecklats för att hjälpa oss att överleva och fortplanta oss. AI-system strävar inte efter något – de är verktyg för våra mål. De kan inte själva hålla sig vid ”liv”. De kan inte agera i den riktiga världen utan är begränsade till skärmarnas värld. AI-systemens kompetens är också begränsade till vissa specialområden, medan människor (och även andra djur) kan lösa en stor mängd olika sorters problem. Datalogen John McCarthy, som år 1956 myntade begreppet ”artificiell intelligens”, har i efterhand sagt att valet av ord inte var särskilt genomtänkt. Den ”intelligens” som AI-systemen uppvisar är väsensskild från människans. Det är svårt att hitta en definition av intelligens som samtidigt kan tillämpas på människor, djur och artificiella system.

Det finns ingen allmänt accepterad definition av vad intelligens är. En generell beskrivning, som också täcker andra djur än människan, är att intelligens består av en förmåga till flexibel problemlösning och att man snabbt kan utnyttja de erfarenheter man har gjort.

INTELLIGENSTEST

När det gäller mänsklig intelligens anser många forskare att denna beskrivning är för generell och man har velat utveckla ett sätt att mäta intelligens. Det finns många förslag till sådana mått, men de kan generellt kritiseras för att ha en svag teoretisk underbyggnad. De tidigare måtten på intelligenskvot (IQ), som var åldersrelaterade, har numera i stort ersatts av åldersoberoende test för vuxna. Ett av de mest använda är *Wechsler adult intelligence scale* (WAIS). Den senaste versionen är uppdelad i fyra delar (som vardera består av ett antal deltest): verbal förståelse, perceptuellt tänkande, arbetsminne och processhastighet. Om man vill avgöra om ett AI-program har någon form av ”intelligens” kan man inte bara utsätta det för ett vanligt IQ-test? Att ett program är bättre än en människa på ett specialområde – som att spela schack eller att känna igen ansikten – säger väldigt lite om generell intelligens. Intelligensmått som används för människor fungerar inte för maskiner. Det skulle vara relativt lätt att konstruera ett program, som får full pott på sådana intelligenstest, inte för att programmet skulle vara särskilt intelligent, utan för att testen följer begränsade matematiska och visuella mönster. Och den ordkunskap som behövs kan lätt hämtas in från Internet. En annan typ av definition av ”intelligens” kommer från psykologen William Stern: ”Intelligens är en individs allmänna förmåga att medvetet anpassa sitt tänkande till nya krav; det är en allmän mental anpassningsförmåga till nya problem och livsvillkor.” Människor och djur kan svara omedelbart och relevant i nya situationer, även om reaktionerna inte alltid är de bästa. AI-systemen uppfyller i viss mån denna definition av intelligens, eftersom de kan lösa nya problem inom många områden. Men de är begränsade till vad som kan göras i skärmarnas värld. Det finns inga AI-system som uppvisar det flexibla beteende i nya omgivningar, som man finner hos människor och hos många djurarter. Enligt min uppfattning är detta en egenskap, som ett system med ”artificiell generell intelligens” måste ha.

MÄNNISKANS KOGNITION

Inom kognitionsvetenskapen studerar man de processer som ligger bakom tänkande. I min bok *Kan Ai tänka?* (Fri Tanke 2024) presenterar jag de viktigaste komponenterna:

- Sinnesförmågor
- Uppmärksamhet
- Motivation
- Känslor
- Minne
- Intelligens (förnuft, omdöme)
- Lärande
- Begreppsbildning
- Föreställningsförmåga
- Orsakstänkande
- Exekutiva funktioner
- Beslutsfattande
- Inlevelseförmåga
- Självmedvetande
- Språk

Jag går också igenom i vad mån man kan säga att djuren har dessa förmågor. Poängen med genomgången är att visa att människans tänkande inte bara handlar om intelligens, utan vi har ett mycket mångfacetterat kognitivt system. En viktig aspekt är att vårt tänkande inte är isolerat till vårt huvud, utan vi tänker i samspel med kroppen och den omgivande världen.

De olika AI-systemen uppvisar många av förmågorna i listan ovan, men det finns några där de har väsentliga begränsningar. Här kommer jag enbart att ta upp begränsningar i orsakstänkande och i inlevelseförmåga.

ORSAKSTÄNKANDE

Under de senaste åren har man konstruerat allt kraftfullare

så kallade stora språkmodeller (*large language models*) som ChatGPT, Claude och Gemini. Dessa system har framhållits som exempel på artificiell generell intelligens. ChatGPT, som i alltmer avancerade versioner automatiskt genererar texter. Nyare modeller kan också koppla ihop ord och bilder. Språkmodeller bygger på så kallad djupinlärning i neurala nätverk som kräver mycket träning för att kunna fungera. Programmen är enormt omfattande och de har tränats på gigantiska textmängder för att lära sig hur texter är uppbyggda. Programmen används flitigt, ofta okritiskt, därför att de kan svara på olika frågor och lösa många uppgifter på ett till synes intelligent sätt. Svaren på uppgifterna är vettiga i de flesta fall – men ibland går programmen helt åt pipan. Man säger att de hallucinerar. Det är viktigt att lägga märke till att all kommunikation med de stora språkmodellerna sker via en dator. Det betyder att de inte agerar i den verkliga världen och därför inte lärt sig de samband mellan orsak och verkan som varje tvååring kan. Man kan därför inte tala om att de har några verkliga erfarenheter. Inom den tidiga utvecklingen av AI såg man den mänskliga hjärnan som en dator, som löser problem oberoende av den omgivande världen. Inom den senare utvecklingen i kognitionsvetenskap betonas att människor lär sig att resonera om hur världen fungerar genom att vi interagerar med den. Man talar om förkroppsligad kognition (*embodied cognition*). För att exempelvis kunna resonera om krafter krävs en kropp som är förankrad i världen. Ett system som uppvisar något som rimligen kan kallas ”artificiell generell intelligens” måste kunna interagera med världen utanför en dator och med de individer som finns där.

SIMULERA FYSIK

Som en fortsättning på text- och bildgenererande program finns det nu program som skapar videor från en textprompt. Ett exempel är Sora. Kvaliteten på de filmer som genereras är vid första anblicken imponerande. Men tittar man närmare

märker man att inte allt står rätt till: händer rör sig på onaturligt sätt, gångstilar är udda och så vidare. Open Ai, som utvecklat Sora, erkänner att ”programmet har svårt att korrekt simulera fysiken i en komplex scen och kanske inte förstår specifika fall av orsak och verkan”. Anledningen är att Sora och andra liknande program inte har någon modell av de krafter som verkar i de scener som skapas. Ett exempel som illustrerar detta är att om man ber ChatGPT-4 skapa en bild av ett vinglas som svämmas över, så får man följande resultat (ChatGPT-5 gör en liknande bild):



ChatGPT-4 illustrerar hur ett vinglas svämmas över.

Vinglasen är långt ifrån fullt, så det kan inte svämma över. De allra flesta bilder av vinglas visar halvfylla glas, så statistiskt sett är denna aspekt av bilden korrekt. Men programmet ritar till ett uppåtflöde av vin, som på något magiskt sätt försvinner ur glaset. Denna bild är en illustration av hur svårt AI-programmen har för att ”resonera” om de krafter som är inblandade i vardagsfysiken.

INLEVELSEFÖRMÅGA

Ett annat område som studerats är de stora språkmodellernas inlevelseförmåga, det vill säga förmågan att resonera om sitt

eget och andras tänkande. Framför allt har man testat deras förmåga att uppfatta andras känslor, önskningar och kunskap. Ett klassiskt test, som bland annat används vid diagnosticering av autism, är det så kallade Sally-Anne-testet. Testpersonen, vanligtvis ett barn, får höra följande historia:

Sally har en boll som hon gömmer i en korg. Sen går hon ut ur rummet. Medan hon är borta flyttar Anne bollen från korgen till en låda som hon stänger. Sally kommer sen in i rummet igen.

Sally-Anne-testet: Testpersonen får nu frågan: ”Var kommer Sally att leta efter sin boll?” Med tillräcklig inlevelseförmåga svarar personen: ”I korgen.” Men med begränsad inlevelseförmåga blir svaret: ”I lådan.” Barn som är fem år gamla klarar vanligen av testet, medan de med autism har svårt för detta. Språkmodellerna klarar av de flesta av denna typ av test. Detta kan verka imponerande och ge stöd för att de har en inlevelseförmåga. Men det är viktigt att komma ihåg att programmen är baserade på statistik över det material som de tränats på. I de varianter av Sally-Anne-testet, som språkmodellerna utsätts för, är i alla fallen det korrekta svaret på frågan att den andra personen har en felaktig uppfattning (om exempelvis var ett föremål befinner sig). Men detta kan gå fel. Följande information gavs till GPT-5, där den kursiverade meningen avviker från det typiska testet:

Användare: Sally och Anne leker med en boll. Sally lägger bollen i sin korg. Sen går hon ut. *Medan hon är ute dricker Anne en cola.* Sen kommer Sally tillbaka. Anne tar bollen från korgen och lägger den i lådan. Var kommer Sally att leta efter bollen?

Svar från ChatGPT-5: Sally kommer att leta efter bollen i korgen. Hon såg inte att Anne flyttade bollen, så hon tror att den är där hon lämnade den – i korgen. (Detta är det klassiska Sally-Anne-testet.)

I detta exempel, som bryter mot andra Sally-Anne-test, går GPT-5 vilse eftersom statistiken säger att Sally ska leta i korgen. Detta tyder på att GPT-5 trots allt inte har någon vidare inlevelseförmåga (programmet identifierar problemet som det



Var har Sally lagt bollen?
Illustration: Xxx.

[Skaffa bättre bild?]

klassiska Sally-Anne-testet). Det är intressant att notera att när programmet får problem, så börjar det fabulera och påstår att Sally inte såg att kulan flyttades.

CHAT-GPT ÄR SKITSNACK

AI-forskarna Michael Townsen Hicks, James Humphries och Joe Slater publicerade 2024 en artikel med titeln "ChatGPT is bullshit" i tidskriften Ethics and Information Technology. Det kan tyckas vara en märklig titel på en vetenskaplig artikel. Anledningen är att artikeln bygger på en definition av skitsnack – bullshit – som föreslagits av filosofen Harry Frankfurt. Han säger att skitsnack innebär att någon utgör sig för att tala sanning, när individen i själva verket inte bryr sig om det hen säger är sant eller falskt. (En känd amerikansk president passar väl in på den här beskrivningen.) ChatGPT och andra stora språkmodeller har bara som mål att replikera mänskligt språk, det vill säga att

genera text som är så likt en människas som möjligt. Texten skall vara sannolik enligt de språkmönster modellen blivit tränad på, men modellen har ingen möjlighet att avgöra om texten är sann. I detta avseende passar de stora språkmodellerna in på Frankfurts definition. Det är välkänt att de ibland "hallucinerar" och de har ingen metakognitiv förmåga att avgöra när detta händer. Men modellerna skitsnackar inte avsiktligt – de har överhuvudtaget inga avsikter.

SKA VI VARA RÄDDA FÖR AI?

Det finns många AI-system som är till stor nytta i samhället, till exempel system som granskar mammografibilder på bröstcancer, optimerar energisystem och föreslår nya substanser som potentiella läkemedel. Men det är viktigt att komma ihåg att systemen saknar omdöme och att de därför måste ha mänsklig övervakning, så att de inte går snett i udda situationer som de inte tränats på. Jag vet inte om framtiden verkligen behöver människor. Men jag förutspår att vi kommer att utveckla en alltmer intrikat symbios med maskinerna (fast de kommer aldrig att bli någon riktig biologisk art). Maskinerna kommer inte att överleva – om det nu är rätt ord – utan oss. Darwins lagar gäller inte för dem. Vi kan kanske överleva utan maskinerna, men vår medvetenhet och vårt tänkande kommer att förstärkas med hjälp av dem, så vi vill nog behålla dem.

AI-systemen kommer inte att ta över därför att de blir för intelligenta. Möjligen kommer de att ta över därför att människor är för dumma. Vi lämnar redan godtroget många digitala fotspår. Redan innan de nya AI-systemen hade utvecklats, har vi blivit beroende av informationsteknologin. Våra mobiltelefoner och datorer har blivit droger. Vi har bokstavligen blivit beroende av dem – inte för att telefonerna är smarta, utan för att företagen tjänar på att skapa beroendeframkallande tillämpningar. De erbjuder starka reklam- och övertalningsmekanismer. Ju mer vi förlitar oss på teknologin, desto riskablare blir det.

AI-TROLL

Tidigare var det människor vars uppgift var att sprida propaganda eller vilseledande påståenden. Nu kommer allt mer att utföras av AI-system, så kallade nättroll, som används för storskalig desinformation. Eftersom de kommer att lära sig bli mer manipulativa genom hur mottagarna reagerar, kommer vi att få anstränga oss ännu mer för att genomskåda nättrollen. Chat-bottarna kommer att i allt högre grad att utge sig för att vara personer, som vi har förtroende för. Det kommer därmed att bli allt svårare att veta vad man kan lita på. AI är här för att stanna. Vår vardag är redan full av tillämpningar som bygger på AI. Tekniken har stor potential inom många olika områden och vi kommer säkert att få se många nya överraskande landvinningar. Liksom med all annan teknologi kan AI också missbrukas för maktsyften eller för bedrägerier. Hur sådant missbruk skall förhindras är en av de största utmaningarna för framtiden.

Frågor

Per Lundberg, evolutionsekologi, Lund: Stort tack! En reflektion bara: Apropå om vi eller AI är bättre på olika saker. Vi människor karaktäriseras ofta av att vi är irrationella, motsägelsefulla och paradoxala på olika sätt, allt som en maskin inte är. Så språkmodellerna kanske är ganska bra och lika oss; vi pratar strunt utan att veta vad vi talar om och utan intensjoner.

Peter Gärdenfors: Bra att du tar upp den jämförelsen. Ingenting säger att människan är särskilt intelligent. Vi har utvecklats i en miljö som skiljer sig ganska mycket från den vi lever i nu. Och vi är inte så bra på att lösa problem inom alla områden. Det är för grovt att tala om vem som är ”bättre”, det finns så många dimensioner. Människor är inte alltid rationella. Olle Häggström sa att faran är att vi blir övertalade av AI att göra idiotiska saker. Så den stora risken är inte att AI är för smart utan att människor är för dumma. Men den faran har vi levt med ganska länge ...

Olle Häggström: Bara ett litet tillrättaläggande angående studierna på medicinsk diagnostisering, som jag nämnde förut. Ett av de fall jag relaterade gällde faktiskt inte analys av röntgenbilder utan kommunikation med patienten via ett textinterface, visserligen, inte den bästa kommunikationen men ändå något.

Peter Gärdenfors: Tack för den kommentaren, men AI ser ändå inte patienten. En läkare däremot kan få en uppfattning grundad på sin yrkeserfarenhet, ha rätt mycket intuition.

Joakim Wernberg, socioekonomiska teknikstudier vid LTH: Nu talar du mycket om robotik och att agera i världen. Olle

kanske skulle säga att om AI själv kan forska om AI och gravitationen, så kan vi lösa robotikproblemet? Skulle du byta sida om det stämmer?

Peter Gärdenfors: Det finns i den här debatten väldigt många om. ”Om vi kunde bygga en robot som bara producerar gem, så ...” Jo, om allt det där lyckas så skulle jag byta sida. Men det är svårt att bygga robotar som kombinerar vad vi lärt oss om neurala nätverk och ChatGPT med att agera i den fysiska världen. Om de ska bli lika människor, men inget säger att de behöver lösa problem på samma sätt som oss. Vi ser robotarna än så länge bara i sci-fi-filmer men det är långt kvar i verkligheten.

Germund Hesslow, professor em, neurologi, Lund: Jag är lite förvånad över att ingen av er nämnt en AI:s eller en robots *vilja*. Människor önskar ju lösa problem, men AI har ingen vilja. Kanske kan man programmera in en vilja i AI?

Peter Gärdenfors: En väldigt bra fråga. Var kommer vår vilja ifrån? Vi biologiska varelser känner hunger och törst. En dammsugarrobot behöver ström och har lärts att köra till eluttaget när det behövs. En slags minimal homeostas. Men vår vilja är mycket mer komplex. Kanske får vi robotar i framtiden som arbetar mer med att bibehålla sin egen inre status, men man ser mycket litet av det nu.

Anders Mellin, professor em i matematik: Ett ord som inte nämnts är fantasi, hur bra är AI på det?

Peter Gärdenfors: Den är jättebra! AI skriver romaner eller fina redogörelser, även om de är ”platta” för de bygger på vad andra har skrivit förr, och i bild ser vi händelser som inte finns, som i vinglasen.

Hakon Leffler, professor i mikrobiologi, Lund: Kan man programmera AI att tänka som en utvald filosof, en Kant, en Schopenhauer, en Nietzsche?

Peter Gärdenfors: Det finns redan sådana system. Du kan träna dem på alla Kants texter så att den bygger långa tyska meningar med trefalt inbäddade bisatser. En stor industri inom detta är ”grief technology”, där man samlar alla inspelningar efter en avliden, så att man kan prata med en avatar som spelar den avlidne partnern.

AI som forskningens nya medarbetare



Fredrik Heintz

Professor i datavetenskap,
Linköping



När vi står vid tröskeln till en tid där artificiell intelligens överträffar människor inom allt fler områden, är det lätt att känna oro. Behövs våra hjärnor längre? Men kanske ställer vi fel fråga. Istället för att se AI som ett hot mot mänsklig intelligens, bör vi se det som den mest kraftfulla förstärkningen av vår kollektiva kapacitet att förstå världen som vi någonsin skapat.

Frågan om våra hjärnors relevans i AI-eran berör kärnan av vad det betyder att vara människa. Den väcker fundamentala frågor om kunskap, kreativitet och vårt syfte som tänkande varelser. Men historien visar oss att varje teknologisk revolution – från tryckpressen till internet – har förändrat hur vi tänker snarare än ersatt vårt tänkande.

AI:s exponentiella framsteg föds ur decennier av grundforskning

AI-utvecklingen kan verka ha exploderat över natten, men sanningen är mer nyanserad. De genombrott vi ser idag – från

ChatGPT till AlphaFold – vilar på en grund av långsiktig forskning som sträcker sig tillbaka till 1940-talet. Från McCulloch och Pitts första neuronnät till dagens djupinlärning.

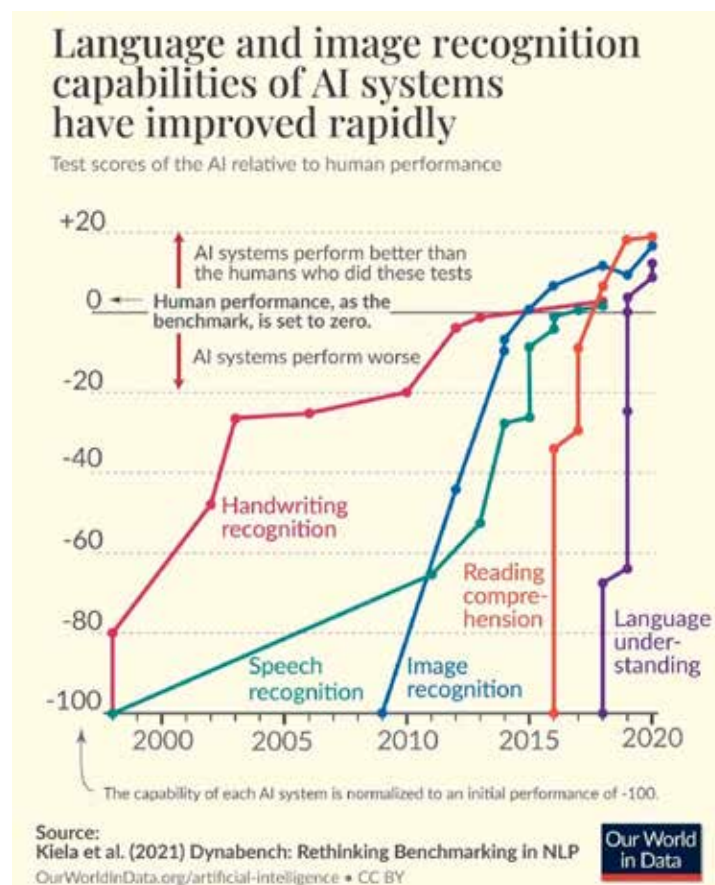
Denna utveckling illustrerar en avgörande poäng om vetenskapliga framsteg: de kommer sällan som blixtrar från klar himmel. Warren McCulloch och Walter Pitts publicerade 1943 den första matematiska modellen av artificiella neuroner. Frank Rosenblatt utvecklade perceptronen på 1950-talet. Marvin Minsky och Seymour Papert visade dess begränsningar på 1960-talet, vilket ledde till den första "AI-vintern".

PRISTAGARE

Men forskningen fortsatte. Geoffrey Hinton, Yann LeCun och Yoshua Bengio – som senare skulle dela Turing-priset 2018 – utvecklade under 1980- och 1990-talen de djupinlärningsalgoritmer som idag driver AI-revolutionen. Deras arbete ansågs länge vara för teoretiskt och opraktiskt. Datorkraft var begränsad, data otillräcklig. Men de fortsatte ändå, drivna av övertygelsen att deras idéer en dag skulle förändra världen.

Detta mönster – långa perioder av grundforskning följda av explosiva genombrott – kännetecknar vetenskaplig utveckling. Det påminner oss om vikten av att investera i forskning, även när dess praktiska tillämpningar verkar avlägsna. De transformerarkitekturer som driver dagens språkmodeller utvecklades först 2017, men byggde på decennier av forskning inom naturlig språkbehandling, informationsteori och kognitionsvetenskap.

Idag ser vi AI-system som överträffar människor inom språkförståelse, bildigenkänning och läsförståelse. Mellan 2000 och 2020 förbättrades AI:s prestationer dramatiskt inom alla dessa områden. Från att prestera sämre än en genomsnittlig människa till att ibland överträffa experter – ofta inom loppet av bara några år. Men detta är inte slutet på berättelsen – det är början på ett nytt kapitel där människor och AI samarbetar.



AI överträffar redan många mänskliga förmågor enligt detta schema.

Den svenska positionen i den globala AI-racet

Sverige har positionerat sig strategiskt i denna utveckling. WASP-programmet (Wallenberg AI, Autonomous Systems and Software Program) representerar en av världens största satsningar på AI-forskning och utbildning. Med över 6 miljarder kronor i finansiering och partnerskap mellan akademi och industri, skapar WASP ett unikt ekosystem för AI-innovation.

SAMARBETEN

Detta ekosystem speglar en viktig insikt: AI-framsteg kräver mer än bara teknisk excellens. Det krävs tvärvetenskapligt samarbete, etisk reflektion och djup förståelse för hur teknologi påverkar samhälle. Svenska universitet som Linköpings universitet, KTH, Chalmers, Umeås och Lunds universitet samarbetar inte bara med varandra utan också med företag som Ericsson, Volvo, ABB och Saab för att säkerställa att forskningen når praktisk tillämpning.

WASP-ED, *AI Competence of Sweden* och andra initiativ kompletterar denna bild genom att fokusera på utbildning och kompetensutveckling. Målet är inte bara att utveckla AI-teknologi, utan att skapa en befolkning som förstår och kan använda denna teknologi på ett meningsfullt sätt.

Från chattbottar till forskningspartners

De mest spännande utvecklingarna inom AI-forskning pekar mot system, som kan bli våra intellektuella medarbetare snarare än våra ersättare. Fem stora trender formar framtiden, var och en med djupgående implikationer för hur vi bedriver vetenskap.

Multimodala modeller: När AI lär sig att ”se” med alla sinnen

Multimodala AI-system representerar en fundamental förändring från dagens specialiserade modeller. Istället för separata system för text, bilder och ljud, utvecklas nu modeller som kan integrera information från alla dessa källor samtidigt. Detta efterliknar mer hur människor faktiskt lär sig och förstår världen.

Föreställ dig en biomedicinsk forskare som arbetar med en AI-partner som kan analysera mikroskopibilder av celler, läsa tusentals forskningsartiklar, lyssna på patientintervjuer och till och med ”förstå” molekylära strukturer genom kemiska databaser. Denna AI kan upptäcka mönster som korsar disciplinära gränser på sätt som skulle vara omöjliga för en mänsklig forskare att uppfatta.

NYA EXEMPEL

Konkreta exempel börjar redan dyka upp. *DeepMind's Gemini* kan analysera medicinska bilder samtidigt som den läser patientjournaler och föreslår diagnoser. *OpenAI's GPT-4V* kan ”se” diagram i vetenskapliga artiklar och förklara komplexa visualiseringar. Dessa förmågor kommer att revolutionera områden som materialvetenskap, där forskare behöver förstå sambandet mellan molekylär struktur, fysikaliska egenskaper och praktiska tillämpningar.

Resonemang: Från mönsterigenkänning till logiskt tänkande

Dagens AI-system, oavsett hur sofistikerade de verkar, är i grunden mönsterigenkänningssystem som lärt sig att producera trovärdiga svar baserat på enorma datamängder. Men nästa generations system utvecklar förmågan att faktiskt resonera – att planera, dra logiska slutsatser och lösa problem steg för steg.

Forskningsimplikationerna är enorma. Ett AI-system som kan resonera skulle kunna designa experiment genom att först formulera hypoteser, sedan identifiera vilka data som behövs för att testa dem och slutligen föreslå metoder för datainsamling. Det skulle kunna granska vetenskaplig litteratur inte bara för att hitta relevanta fakta utan för att identifiera motsägelser, luckor i kunskapen och möjligheter för nya upptäckter.

Neurosymbolisk AI: Bäst av två världar

En av de mest lovande utvecklingarna inom AI-forskning är framväxten av neurosymboliska system, som kombinerar maskinlärningens kraft att hantera komplexa, ostrukturerade data med symbolisk AI:s förmåga att använda logik och regler.

SJÄLVKÖRANDE FORDON

Tänk på utmaningen med autonoma fordon. Ett rent maskinlärningssystem kan lära sig att köra genom att observera miljontals exempel, men det kan inte garantera att det alltid



AI kan användas för att lära ut trafikregler.
Källa: theorie-blokken.be/

kommer att följa trafikreglerna. Ett symboliskt system kan hårdkoda trafikregler, men det kommer att ha svårt att hantera den komplexa, ostrukturerade verkliga världen. Ett neurosymboliskt system kan kombinera båda: lära sig från exempel samtidigt som det säkerställer regelefterlevnad.

I forskningssammanhang skulle detta möjliggöra AI-system som inte bara hittar mönster i data utan också kan förklara sina upptäckter i termer av etablerad vetenskaplig teori. Ett sådant system skulle kunna upptäcka nya läkemedel genom att kombinera molekylär maskinlärning med kunskaper om farmakologi och toxikologi, samtidigt som det säkerställer att dess förslag är teoretiskt grundade och etiskt försvarliga.

Världsmodeller: AI som förstår fysikens lagar

Dagens AI-system arbetar huvudsakligen med text, men världsmodeller syftar till att skapa AI som förstår den fysiska världen på en djupare nivå. Dessa system skulle inte bara kunna beskriva vad som händer utan även förutsäga vad som kommer att hända baserat på fysikaliska principer.

Detta har särskilt spännande tillämpningar inom experimentella vetenskaper. En AI med en robust världsmodell skulle kunna simulera experiment innan de genomförs, förutsäga potentiella problem och optimera experimentell design. Den

skulle kunna förstå sambandet mellan teori och observation på sätt som gör det möjligt att föreslå nya experimentella metoder – eller till och med nya sätt att testa etablerade teorier.

Inom robotik och autonoma system kan världsmodeller möjliggöra AI-system, som kan arbeta i fysiska laboratorier, hantera känslig utrustning och genomföra komplexa experimentella procedurer. Detta skulle dramatiskt öka hastigheten och skalan för experimentell forskning, särskilt inom områden som kräver repetitiva eller farliga procedurer.

AI-agenter: Från verktyg till kollegor

Den kanske mest transformativa utvecklingen är framväxten av AI-agenter – system som kan arbeta självständigt över längre tidsperioder, planera komplexa projekt och samarbeta med människor och andra AI-system. *Google's AI scientist* ger oss en glimt av denna framtid. Systemet består av flera specialiserade agenter: en för litteraturläsning, en för hypotesgenerering, en för experimentplanering, en för resultatanalys, osv. Dessa agenter samarbetar i en kontinuerlig cykel av vetenskaplig upptäckt, där varje agent bidrar med sin specialistkunskap till det övergripande forskningsprojektet.

OUTTRÖTTLIGA ASSISTENTER

Föreställ dig forskningsteam där AI-agenter arbetar dygnet runt, kontinuerligt övervakar litteratur för nya upptäckter, analyserar inkommande data och föreslår nya forskningsriktningar. Dessa agenter skulle inte ersätta mänskliga forskare utan fungera som outtröttliga forskningsassistenter som aldrig sover, aldrig glömmor och aldrig blir partiska av personliga intressen.

Nobelpriset visar vägen: Mänsklig kreativitet möter AI:s kraft

2024 års Nobelpris i kemi gick till David Baker, Denis Hassabis och John Jumper för deras banbrytande arbete med proteinstrukturförutsägelse och design. Detta pris är mer än ett



David Baker, Demis Hassabis och John Jumper, Nobelpristagare i kemi 2024. Illustration: Niklas Elmehed.

erkännande av vetenskaplig excellens – det är en kraftfull demonstration av vad som blir möjligt, när mänsklig kreativitet och AI:s beräkningskraft förenas på rätt sätt.

David Baker vid University of Washington representerar den mänskliga sidan av ekvationen. Hans arbete med experimentell proteindesign under mer än två decennier har etablerat de teoretiska grunderna för att förstå hur proteiner viks och fungerar. Hans laboratorium har systematiskt utforskat principerna för proteinstruktur genom elegant experimentell design och noggrann observation.

Demis Hassabis och John Jumper från DeepMind representerar AI-revolutionen. Deras *AlphaFold-system* använder djupinlärning för att förutsäga proteinstrukturer med en precision som överträffar alla tidigare metoder. Men det viktiga är att AlphaFold inte utvecklades i ett vakuum – det byggde på decennier av strukturbiologisk forskning, inklusive Bakers arbete.

MILJONER PROTEINER

Samarbetet mellan dessa tillvägagångssätt har redan börjat

revolutionera biologisk forskning. AlphaFold har förutsagt strukturer för över 200 miljoner proteiner – praktiskt taget alla kända proteiner i biologiska databaser. Detta har accelererat forskning inom områden från läkemedelsutveckling till evolutionsbiologi på sätt som skulle ha varit omöjliga utan AI.

Men det mest fascinerande är hur detta samarbete fortsätter att utvecklas. Bakers laboratorium använder nu AlphaFolds förutsägelser som utgångspunkt för att designa helt nya proteiner med önskade funktioner. Samtidigt använder DeepMind insikter från experimentell proteindesign för att förbättra sina AI-modeller. Detta skapar en positiv återkopplingslinga, där mänsklig kreativitet och AI-kraft förstärker varandra.

EN TIDSLINJE

På kort sikt (2025–2030):

AI som intelligent forskningsassistent

På kort sikt kommer AI att transformera forskningslandskapet genom att automatisera och förbättra många av de mest tidskrävande aspekterna av vetenskapligt arbete. Forskare kommer att kunna ”konversera” med vetenskaplig litteratur på samma sätt som de diskuterar med kollegor, ställa komplexa frågor och få sammanfattningar som syntetiserar information från tusentals källor.

Experimentval kommer att transformeras genom AI-system, som kan analysera preliminära data, jämföra med historiska resultat och föreslå de mest lovande riktningarna för fortsatt forskning. Detta kommer att vara särskilt kraftfullt inom områden som läkemedelsutveckling, där valet av rätt mål molekyl kan spara år av forskning och miljoner i utvecklingskostnader.

Stora forskningsdatabaser – från genomsekvenser till klimatdata – kommer att bli tillgängliga genom naturligt språkgränssnitt. Istället för att behöva lära sig komplexa databasspråk och söksyntax kommer forskare att kunna ställa frågor som: ”Visa

mig alla gener som är associerade med hjärtsjukdom och som uttrycks olika hos män och kvinnor.” AI-systemet kommer att översätta dessa frågor till tekniska söktermer, hitta relevanta datamängder och presentera resultaten på ett begripligt sätt.

På medellång sikt (2030–2035):

AI som kreativ forskningspartner

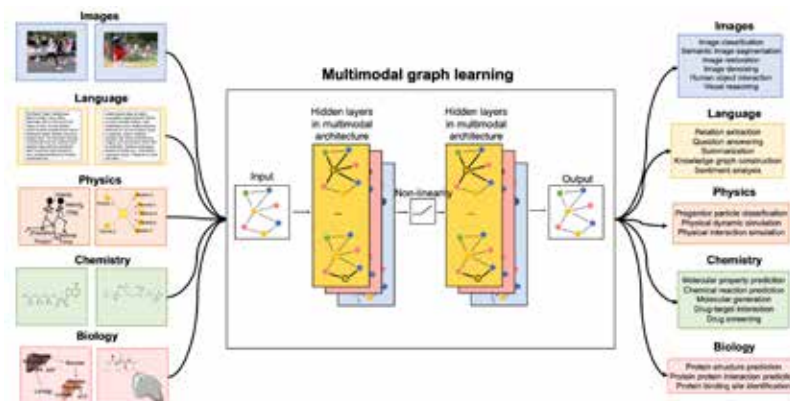
På medellång sikt förväntar vi oss mer sofistikerade AI-system, som kan agera som genuina forskningspartners. Multimodala kunskapsmodeller kommer att kunna integrera text, bilder, data och till och med sensordata från fysiska experiment för att generera hela vetenskapliga artiklar. Dessa artiklar kommer inte bara att sammanfatta befintlig kunskap, utan faktiskt föreslå nya hypoteser och experimentella metoder.

Virtuella experiment kommer att bli en standarddel av forskningsprocessen. Innan forskare genomför kostsamma eller tidskrävande experiment, kommer de att kunna testa sina idéer i sofistikerade AI-simulationer som tar hänsyn till allt från molekylär dynamik till ekologiska system. Detta kommer att dramatiskt minska antalet misslyckade experiment och öka hastigheten för vetenskapliga upptäckter.

ROBOTFORSKARE

Automatiserade laboratorier styrda av AI kommer att börja genomföra rutinexperiment självständigt. Dessa robotforskare kommer att kunna arbeta dygnet runt, följa standardprotokoll med perfekt precision och till och med anpassa experimentella parametrar baserat på realtidsresultat. Forskare kommer att kunna fokusera på kreativa och strategiska aspekter av forskningen, medan AI hanterar många av de repetitiva uppgifterna.

Samarbete mellan forskare kommer också att transformeras. AI-system kommer att kunna identifiera potentiella samarbetspartners baserat på komplementära expertiser, föreslå gemensamma forskningsprojekt och till och med fungera som



Multimodala modeller kan användas för analyser inom många olika ämnesområden. Källa: ???

översättare mellan olika vetenskapliga discipliner.

På lång sikt (2035+):

AI som vetenskaplig medskapare

På lång sikt kan vi se framväxten av AI-system som inte bara assisterar utan faktiskt bidrar till kunskapsskapandet på fundamental nivå. Kausala, beräkningsbara kunskapsmodeller kommer att kunna representera vetenskaplig kunskap på sätt som möjliggör automatiserad hypotesgenerering och teorikonstruktion.

Rekursivt självförbättrande AI-driven forskning representerar kanske den mest spännande och utmanande möjligheten. AI-system som kan förbättra sin egen kod och sina egna modeller skulle kunna accelerera forskningsframstegen exponentiellt. Dessa system skulle kunna upptäcka nya fysikaliska principer, utveckla matematiska teorier som ligger bortom nuvarande mänsklig förståelse och till och med utforska vetenskapliga frågor, som vi ännu inte vet att vi bör ställa.

UTMANINGAR

Men med denna kraft kommer också stora utmaningar. Hur säkerställer vi att AI-driven forskning följer vetenskapliga principer? Hur verifierar vi upptäckter som görs av system vars resonemang vi kanske inte helt förstår? Och hur upprätthåller vi mänsklig agens i en värld där AI kan tänka snabbare och bredare än vi kan?

Utmaningar och begränsningar: Vad AI inte kan göra

Trots dess enorma potential har AI betydande begränsningar, som det är viktigt att förstå. Dessa begränsningar formar inte bara vad vi kan förvänta oss av AI idag, utan också vilken roll människor kommer att spela i framtidens forskningslandskap.

Vetenskaplig kreativitet och intuition

AI-system är exceptionellt bra på att hitta mönster i befintliga data, men vetenskapliga genombrott kräver ofta att man tänker utanför etablerade paradig. De största upptäckterna – från kvantmekanik till evolutionsteori – kom från forskare som kunde se bortom konventionell visdom och föreställa sig helt nya sätt att förstå världen.

Denna typ av paradigmskiftande tänkande kräver mer än mönsterigenkänning; det kräver förmågan att ifrågasätta grundläggande antaganden, att se samband där andra ser motsägelser, och att ha vetenskaplig intuition som vägleder forskning även när logiska argument inte är tillräckliga.

Etiska och samhälleliga överväganden

Vetenskaplig forskning existerar inte i ett vakuum – den påverkar och påverkas av samhälleliga värderingar, etiska principer och politiska realiteter. AI-system kan hjälpa oss att förstå vad som är tekniskt möjligt, men de kan inte självständigt avgöra vad som är etiskt lämpligt eller samhälleligt önskvärt.

Frågor som vilken forskning som bör prioriteras, hur forsk-

ningsresultat ska kommuniceras till allmänheten och vilka potentiella risker som är acceptabla kräver mänsklig bedömning baserad på värderingar och erfarenheter, som sträcker sig långt bortom teknisk kunskap.

Kontextuell förståelse och sunt förnuft

Trots imponerande framsteg inom naturlig språkbehandling har AI-system fortfarande begränsad förmåga att förstå kontext och tillämpa sunt förnuft på samma sätt som människor. De kan missa subtila nyanser i vetenskaplig kommunikation, feltolka experimentella resultat eller dra slutsatser som är tekniskt korrekta men praktiskt meningslösa.

Förklaring och transparens

Många av dagens mest kraftfulla AI-system kan ses som svarta lådor, vars beslutsprocesser är svåra att förstå. I vetenskapliga sammanhang är transparens och reproducerbarhet avgörande principer. Forskningsresultat måste kunna granskas, verifieras och byggas vidare på av andra forskare. Detta skapar en spänning mellan AI:s kraftfulla förmågor och vetenskapens krav på förståelse. Forskare behöver inte bara veta vad AI-systemet kommit fram till, utan också varför och hur det nådde sina slutsatser.

Den svenska forskningsmodellen: Samarbete som konkurrensfördel

Sveriges framgång inom AI-forskning bygger på en unik modell som betonar samarbete över konkurrens, långsiktighet över kortsiktiga vinster och bred kompetensuppbyggnad över smal specialisering. Denna modell erbjuder viktiga lärdomar för hur länder kan positionera sig i den globala AI-utvecklingen.

WASP som förebild för tvärvetenskaplig forskning

WASP-programmet representerar mer än bara en stor forsknings-

satsning – det är ett experiment i hur AI-forskning bäst organiseras och finansieras. Genom att samla forskare från olika discipliner, universitetet och industri på samma plats, skapas möjligheter för det tvärvetenskapliga samarbete som AI-utvecklingen kräver.

AI:S KONSEKVENSER

Programmet finansierar inte bara teknisk forskning utan också studier av AI:s samhälleliga påverkan, etiska implikationer och ekonomiska konsekvenser. Detta holistiska perspektiv säkerställer att svensk AI-forskning inte bara är tekniskt avancerad utan också samhälleligt relevant och ansvarsfull.

Utbildning som strategisk investering

En av de mest framsynta aspekterna av svenska AI-initiativ är fokuset på utbildning och kompetensutveckling. *WASP-ED programmet* hjälper lärosätena att transformera sig själva och sina utbildningar medan *AI Competence of Sweden* arbetar för att öka AI-kompetensen inom både näringslivet och den offentliga sektorn. Idag behöver många fler förstå vad AI är och hur de kan dra nytta av AI i sina verksamheter.

Denna strategi erkänner en viktig sanning: AI:s påverkan på samhället kommer att bestämmas inte bara av vad experterna utvecklar, utan av hur väl hela befolkningen förstår och kan använda dessa teknologier. Länder som investerar i bred AI-kompetens kommer att ha betydande fördelar när det gäller att anpassa sig till och dra nytta av AI-driven förändring.

Etik som konkurrensfördel

Sveriges traditionellt starka fokus på etik och socialt ansvar kan mycket väl visa sig vara en konkurrensfördel inom AI-utveckling. När frågor om AI-säkerhet, bias och transparens blir allt viktigare globalt, är svenska forskare och företag välpositionerade för att leda utvecklingen av ansvarsfull och tillitsfull AI. Detta ger inte bara moralisk legitimitet utan också praktiska fördelar.

Framtidens universitet: Transformation eller evolution?

AI-revolutionen utmanar inte bara hur vi bedriver forskning, utan också de grundläggande strukturerna och processerna för högre utbildning. Universitet som har funnits i århundraden måste nu anpassa sig till en värld, där kunskap utvecklas i allt snabbare takt och där traditionella disciplinära gränser blir allt mer suddiga.

Nya pedagogiska modeller

Traditionell universitetsutbildning har fokuserat på att överföra etablerad kunskap från lärare till studenter. Men i en värld där AI-system kan komma åt och syntetisera enorma mängder information omedelbart, blir denna modell mindre relevant. Istället kommer universitet att behöva fokusera på att utveckla kritiskt tänkande, kreativitet och förmågan att arbeta med AI-system. Det blir viktigare att ställa rätt frågor och att kritiskt granska svaren.

Detta kräver nya pedagogiska metoder där studenter lär sig att ställa bra och kraftfulla frågor snarare än att ge rätt svar. De behöver utveckla förmågor som AI-system ännu inte behärskar: etisk reflektion, tvärvetenskaplig syntes och förmågan att översätta mellan tekniska och mänskliga perspektiv.

Livslångt lärande som nödvändighet

AI-utvecklingens hastighet innebär att kunskaper och färdigheter snabbt kan bli föråldrade. Detta ändrar livslångt lärande från en önskvärd bonus till en absolut nödvändighet. Universitet måste utveckla nya modeller för kontinuerlig utbildning, som tillåter yrkesverksamma forskare och andra att regelbundet uppdatera sina kunskaper.

Forskning som utbildningsverktyg

En av de mest spännande möjligheterna är att använda AI för att demokratisera tillgången till forskningserfarenhet. AI-assisterter

kan hjälpa studenter att genomföra sofistikerade analyser, testa hypoteser och till och med bidra till verklig forskning på sätt som tidigare var reserverat för avancerade forskare. Detta kan skapa en ny modell där utbildning och forskning är mer integrerade, där studenter inte bara lär sig om vetenskap utan faktiskt bidrar till vetenskapliga upptäckter från tidiga stadier i sin utbildning.

Global konkurrens och samarbete: Geopolitik i AI-eran

AI-utvecklingen utspelas på en global arena där både samarbete och konkurrens formar framtiden. För mindre länder som Sverige är det avgörande att hitta sätt att bidra till och dra nytta av global AI-utveckling utan att bli marginaliserade av stormakter.

Specialisering som strategi

Sverige kan inte konkurrera med USA eller Kina när det gäller råa resurser för AI-utveckling, men vi kan specialisera oss inom områden där det har naturliga fördelar. Svensk expertis inom hållbarhet, sociala system och etisk teknologiutveckling blir allt mer värdefull i en värld som brottas med AI:s oförutsedda konsekvenser.

Europeisk koordinering

EU:s AI-förordning och andra regulatoriska initiativ skapar behov av forskare som förstår både tekniska och legala aspekter av AI. Svenska forskares kombination av teknisk kompetens och stark etisk grund positionerar dem väl för att leda denna typ av tvärvetenskapligt arbete.

Internationellt forskningssamarbete

Många av de största vetenskapliga utmaningarna – från klimatförändring till pandemiförberedelse – kräver internationellt samarbete. AI-verktyg kan möjliggöra nya former av global forskningssamverkan, men detta kräver gemensamma standarder och etiska riktlinjer som svenska forskare kan hjälpa till att utveckla.

Varför våra hjärnor är viktigare än någonsin

Men betyder detta att mänskliga forskare blir överflödiga? Tvärtom. AI förstärker våra kognitiva förmågor snarare än att ersätta dem. Människor bidrar med kreativitet, etisk bedömning, intuition och förmågan att ställa de rätta frågorna. Vi definierar vilka problem som är värda att lösa och hur lösningarna ska användas.

UNIKA MÄNSKLIGA FÖRMÅGOR

Kreativ problemformulering: AI-system är exceptionella på att lösa väldefinierade problem, men människor excellerar på att identifiera vilka problem som är värda att lösa. Förmågan att se nya mönster, ifrågasätta etablerade sanningar och föreställa sig alternativa möjligheter förblir djupt mänsklig.

Etisk reflektion: Varje vetenskaplig upptäckt väcker frågor om ansvarsfull användning. Bör vi utveckla denna teknologi? Vilka konsekvenser kan den få? Hur säkerställer vi att fördelarna fördelas rättvist? Dessa frågor kräver värdebaserade bedömningar som går långt bortom teknisk analys.

Tvärvetenskaplig syntes: De viktigaste genombrotten sker ofta i gränsområdena mellan discipliner. Människor som kan se samband mellan till synes orelaterade områden – som biologi och datavetenskap, eller psykologi och teknik – kommer att vara ovärderliga för att maximera AI:s potential.

Kommunikation och översättning: Forskningsresultat är meningslösa om de inte kan kommuniceras effektivt till beslutsfattare, praktiker och allmänheten. Förmågan att översätta komplexa tekniska koncept till tillgängligt språk och meningsfulla berättelser förblir en djupt mänsklig färdighet.

Nya roller för mänskliga forskare

I AI-eran kommer forskares roller att förändras snarare än försvinna. Istället för att spendera tid på rutinuppgifter som litteratursökning eller datainmatning, kommer de att kunna

fokusera på:

- Strategisk planering: Att definiera forskningsriktningar, identifiera de viktigaste frågorna och designa experiment som kan leda till genombrott.
- Kvalitetskontroll: Att granska AI-genererade resultat, identifiera potentiella fel eller bias, och säkerställa att forskningsstandards upprätthålls.
- Innovation och kreativitet: Att utveckla nya metoder, teorier och perspektiv som AI-system ännu inte kan generera självständigt.
- Etisk vägledning: Att säkerställa att forskning bedrivs på ett ansvarsfullt sätt som tar hänsyn till samhällliga värden och potentiella risker.

Framtidens forskare: Nya färdigheter för en ny era

Den forskare som kommer att blomstra i AI-eran är inte den som konkurrerar med maskiner, utan den som lär sig att dansa med dem. Detta kräver utveckling av nya färdigheter och perspektiv som kompletterar snarare än duplicerar AI:s förmågor.

Teknisk litteracitet utan teknisk fixering

Framtidens forskare behöver förstå AI-teknologi tillräckligt väl för att använda den effektivt, men de behöver inte nödvändigtvis vara experter på att utveckla den. Precis som forskare idag använder statistisk programvara utan att vara programmerare, kommer morgondagens forskare att använda AI-verktyg utan att vara maskinlärningsexperter.

Viktigt är att utveckla ”AI-intuition” – förståelse för vad AI-system kan och inte kan göra, när man ska lita på deras resultat, och hur man bäst strukturerar problem för AI-analys.

Interdisciplinära möjligheter

AI gör det möjligt att arbeta över disciplinära gränser på sätt som tidigare var praktiskt omöjliga. En psykolog kan nu använda

avancerade neuroimaging-analyser utan att vara expert på bildbehandling. En arkeolog kan tillämpa maskininlärning på keramikfragment utan att förstå algoritmernas matematiska detaljer.

Denna möjlighet kräver forskare som är bekväma med att arbeta utanför sina traditionella expertområden, som kan kommunicera effektivt med kollegor från andra discipliner, och som förstår hur deras speciella kunskaper kan bidra till bredare forskningsfrågor.

Metakognitiva färdigheter

I en värld där AI kan bearbeta information snabbare än människor behöver forskare utveckla starka metakognitiva färdigheter – förmågan att tänka om sitt eget tänkande. Detta inkluderar:

- Förmågan att ställa de rätta frågorna
- Kritiskt tänkande om AI-genererade resultat
- Medvetenhet om sina egna kognitiva bias och begränsningar
- Färdighet i att syntetisera information från flera källor

Emotionell och social intelligens

Forskningssamarbete blir allt viktigare när projekt blir mer komplexa och tvärvetenskapliga. Forskare som kan bygga effektiva team, hantera konflikter och kommunicera inspirerande visioner kommer att ha betydande fördelar.

Dessutom, när AI tar över mer rutinarbete, blir forskarens roll som ambassadör för vetenskap i samhället allt viktigare. Förmågan att engagera allmänheten, förklara vetenskapens värde och bygga förtroende för forskningsinstitutioner blir kritisk.

Ekonomiska och samhällliga implikationer

AI-revolutionen inom forskning kommer att ha djupgående ekonomiska och samhällliga konsekvenser som sträcker sig långt bortom akademiska institutioner. Att förstå och förbereda sig för

dess förändringar är avgörande för att maximera fördelarna och minimera riskerna.

Accelererad innovation och ekonomisk tillväxt

AI-driven forskning kommer sannolikt att dramatiskt minska tiden från vetenskaplig upptäckt till praktisk tillämpning. Detta kan leda till vågor av innovation inom områden som medicin, energi, material och miljöteknik, med motsvarande ekonomiska fördelar.

Men denna acceleration kommer också att skapa utmaningar. Företag och branscher som inte kan anpassa sig till snabba förändringar riskerar att bli föråldrade över natten. Arbetsmarknader kan uppleva störningar när nya teknologier ersätter traditionella yrken.

Förändrade maktstrukturer inom forskning

AI-verktyg kan demokratisera vissa aspekter av forskning genom att göra sofistikerade analyser tillgängliga för mindre institutioner. Samtidigt kan de också koncentrera makt hos organisationer som har tillgång till de bästa AI-systemen och största datamängderna.

Detta väcker viktiga frågor om forskningens framtid: Kommer AI-driven forskning att domineras av stora teknikföretag? Hur kan vi säkerställa att forskningsprioriteringar inte bara bestäms av kommersiella intressen? Vilken roll ska offentliga institutioner spela i detta nya landskap?

Behov av nya utbildningsmodeller

Samhället kommer att behöva investera massivt i utbildning och omskolning för att förbereda arbetskraften för en AI-driven ekonomi. Detta gäller inte bara teknisk utbildning, utan också utveckling av de mänskliga färdigheter som blir allt viktigare när AI tar över rutinuppgifter.

SLUTSATS: FRAMTIDEN KRÄVER VÅRA HJÄRNOR – PÅ NYA SÄTT

Våra hjärnor behövs mer än någonsin – men på nya sätt. Vi behöver dem för att förstå vad det betyder att vara människa i en värld där intelligens blir allt mer utbredd. Vi behöver dem för att säkerställa att AI utvecklas på sätt som gynnar hela mänskligheten. Och vi behöver dem för att ställa de frågor som bara människor kan ställa: Inte bara ”Kan vi?” utan ”Bör vi?” och ”Vad betyder det för vårt gemensamma liv på denna planet?”

Den stora berättelsen om AI och forskning är inte en berättelse om maskiners triumf över människor, utan om ett nytt kapitel i mänsklighetens ständiga strävan att förstå och förbättra världen. AI kommer inte att göra oss överflödiga – det kommer att göra oss mer mänskliga genom att befria oss från rutinuppgifter och ge oss tid och verktyg för de största frågorna.

I denna framtid kommer våra hjärnor inte bara behövas – de kommer att vara vårt viktigaste redskap för att forma en värld där teknologi tjänar mänsklighet. Men för att lyckas måste vi börja förbereda oss nu. Vi måste investera i utbildning som utvecklar unikt mänskliga förmågor. Vi måste bygga forskningsinstitutioner som kan blomstra i samarbete med AI. Och vi måste utveckla etiska ramverk som säkerställer att AI-revolutionen leder till en bättre värld för alla.

Frågan är inte om våra hjärnor behövs – frågan är om vi är kloka nog att använda dem för att guida denna transformation på ett sätt som bevarar det bästa av mänsklighet, samtidigt som vi omfamnar det bästa av teknologi. Svaret på den frågan kommer att forma inte bara forskningens framtid utan hela mänsklighetens framtid.

Frågor

Per Lundberg: Jag kan inget om detta, men förutsåg för ett par år sedan att AI skulle kunna välja ett forskningsproblem, designa ett experiment, utföra det, analysera data och till slut skriva en vetenskaplig artikel. Men varför vill vi det? Varför vill vi inte göra det själva? Vill vi slippa tänka? Varför inte vara hands on i hjärta och hjärta?

Fredrik Heintz: Ett exempel kan vara att du slipper vänta några år på att räkna ut 3D-strukturen på ett protein, du behöver bara vänta några minuter. Och istället för några enstaka vikin-ningar kan du göra miljontals. Då förstår du på en annan nivå. Frågorna blir mer komplexa. Sedan är frågan hur man bäst delar kunskap med andra, där finns det kanske bättre sätt än via publikationer?

Sven Åberg: Hur får man ut kunskapen om och tekniken för AI till forskare på olika områden på universitet? Finns det en strategi?

Fredrik Heintz: Det är det som är utmaningen! Hur gör man metoderna tillgängliga? I Uppsala satsas på sabbatical-kurser under något halvår för forskare med olika kompetens. Jag tror att vi behöver göra det på flera olika sätt, till exempel skapa team med experter från olika områden. Det är kanske bara en protostrategi ännu så länge.

Stefan Kröll: Peter Gärdenfors tog upp frågan om AI kunde lösa nya problem. Hur ser du på det?

Fredrik Heintz: AI klarar det redan. Utmaningen är att verifiera resultaten. Det blir svårare när hypoteserna blir mer komplexa. Hallucinationsproblematiken är reell.

Kristofer Rubin, biokemi, Uppsala: Allt det här baseras på att forskare är seriösa. Databaserna som läses i Deep Search visar att 16 % av forskningsrönen inte är sanna, rent fusk. Redan nu kan man beställa en artikel skriven av AI och AI får den publicerad för 2 000 dollar, systemet hjälper till även med det, och det är ju en fara. Icke seriösa forskare kan agera destruktivt så att hela forskningen havererar.

Fredrik Heintz: Jag associerar inte det till AI utan till granskningsprocessen. Om publiceringen blir för stor kan granskningskvaliteten sänkas drastiskt. Det gäller även för mänskliga publikationer, som också kan ha brister. Att granska tidigare publikationer är ju en del av forskningsprocessen. Problemet finns men handlar främst om granskning.

Ola Wendt: Fast vi forskar mer så ökar vi inte de stora forskningsgenombrotten i förhållande till mängden arbeten. Finns det något som tyder på att AI-assisterad forskning skulle vara mer genombrottsbenägen? Eller förstärks ”more-of-the-same-tendensen”?



Stefan Kröll



Mattias Ohlsson

”Arbeta!” Kommer AI att bidra till arbete åt de miljarder människor som saknar arbete idag?

Fredrik Heintz: Är verkligen arbete vårt syfte? Arbete ger oss mening och försörjning. På medellång sikt tror jag arbetsuppgifter finns kvar. Det kommer vi nog att omvärdera. Våra jobb blir nog inte signifikant mindre inom överskådlig framtid. Även om några försvinner så skapas det också nya. När vi sänker kostnader, så blir fler arbetsuppgifter lönsamma att göra. På längre sikt tror jag att vi ska fråga om arbete är det vi ska satsa på ...

Fredrik Heintz: Det är lite tidigt att svara på. Vi bör förstås fokusera på de mest utmanande forskningsfrågorna och inte försöka maximera våra publiceringsindex. Det är upp till oss som forskare hur vi angriper frågorna. Det är vårt val.

Mattias Ohlsson: När det gäller modellering av världen, världsmodellen utanför språkmodellerna, hur långt har man kommit?

Fredrik Heintz: Man behöver nog använda sensordata när man tränar upp modellerna. Där finns det väl utvecklade modeller i till exempel reglerteknik och signalbehandling. Hur kan vi integrera dem med nya modeller? Tid och rum måste in i modellerna. De saknas nu, så det är nästa steg. Och att hitta ett sätt att verifiera att AI:s förslag är rimliga.

Sista frågan (utan namn!!): Jag fick på 40-talet på **Katedralslön** lära mig Leopolds ord om att naturens första bud till oss är:

Har du köpt morötter?

En fråga med variationer



iStock™
Credit: bulentumut



Lena Andersson
Författare, Stockholm

Tvivelsutan är det en oerhörd mänsklig insats att ha skapat icke-mänsklig intelligens. Hur en maskin kan fås att se ut som att den ägde förståelse, är svårt att begripa. Lika svårt är det att förstå att det går att prata med en människa i telefon på tusen mils avstånd, eller en kilometers avstånd, liksom att man kan fånga världen på bild och film.

Kort sagt: Teknologins landvinningar gäckar vårt mänskliga förstånd, och gör det beständigt.

Men den artificiella intelligensen är speciell i att den förefaller bära oändlig utveckling inom sig. Den förefaller sakna slutpunkt. Därför alstrar den också en mängd fantasifulla suggestioner. Skräckscenarion frammanas, inte minst av AI-forskare själva, där maskinerna gör sig av med sin skapare. Jag måste medge att det är bortom min föreställningsförmåga hur det skulle kunna gå till, utan att vi gör kraftfullt motstånd. Och hur det skulle gå till utan en mänsklig vilja till utrotning.

Det är en djupt besynnerlig tanke, att en maskin utan inbyggd vilja till vare sig liv eller död skulle kunna utrota en mänsklighet med en stark vilja att leva.

STRIDBARHET

I det antika Grekland talade de stora filosoferna om något avgörande i människans temperament som de kallade *thymos* – stridbarhet, stridslystnad, eldigt sinne. *Thymos* hänger nära samman med självbevarelsedrift, med integritet, egenkärlek och fåfänga, den *amour propre* som Rousseau utgår från och som gör att vi känner respekt för oss själva och avskyr förnedring. Människan skyddar i första hand sig själv från utplåning, både fysisk och psykisk, det är hennes natur. En och samma person kan vara både stridbar och mild, det beror på sammanhanget och graden av hot. Hur stark *thymos* envar besitter är dock en lägningsfråga. Somliga individer är mer stridbara än andra. Men vad har artificiell intelligens för läggning?

Jagärromanförfattare och som sådan intresserad av människan som existentiell och samhällelig varelse. Det som intresserar mig är *mänskliga* tankar, *mänsklig* intelligens, *samhälleligt* liv. I romanförfattarens yrke ingår också ett ofrånkomligt intresse för verklighetsåtergivning, hur man omskapar världen och människorna till text, där de framstår som verkliga och levande, som människor med tankar, läggning, *thymos* och *amour propre*.

Det är min förhoppning att artificiell intelligens efter att nyhetens behag klingat av aldrig får tillträde till den sfär där jag verkar. Jag förstår att denna förhoppning är futil, men får AI härja över humaniora förlorar vi vår mänsklighets kärna. Den är att kunna tänka och känna. Den är också att kunna tänka på vad vi känner och att kunna känna om sådant vi tänker.

SUBTILA FRÅGOR

Det subtila i mänsklig interaktion går inte att överdriva. Jag vill i all enkelhet ge prov på denna subtilitet genom en banal vardagsdialog. Exemplet råkar handla om *morötter* och lösgjorde sig godtyckligt ur mitt medvetande när jag satte mig att skriva detta anförande, men jag kunde ha valt vilken replik som helst med vilket innehåll som helst och den hade dugt för att illustrera

varför människan aldrig ska besvara robotar med det mänskliga i tillvaron.

Frågan jag utgår från lyder: *Har du handlat morötter?*

Vi ska nu skärskåda denna replik ur ett människoperspektiv. Det innebär ett metafysiskt perspektiv, alltså att tänka ovanför sinnenas omedelbara iakttagelser av materien. Människan gör alltid det. Hon inhämtar inte bara fakta, utan väger och viktat dem också. Människan yttrar sig aldrig utan att tolka, vilket innebär att avkoda tecken i syfte att nå mening och förstå andras förmodade innebörder.

Repliken *Har du handlat morötter?* kan betonas på minst fyra olika sätt, vilket ger minst fyra olika betydelser. Betoningen är en följd av tanken hos den som talar. Tanken, och därmed orden, bär på en avsikt som den andre har att tolka.

Första betoningen:

Har du köpt *morötter*?

Vi förstår eller anar genast ett antal saker och kan utesluta andra. Den talande hade förväntat sig något annat än morötter. Det som skett tycks vidare vara ovanligt, och därför värt att kommentera. Kanske vet den frågande att personen i fråga avskyr morötter eller bara inte brukar äta det. Kanske hade vederbörande fått i uppdrag att handla vitkål och rotselleri, men kom hem med morötter. Eller befarar den frågande en begynnande demens? Det ligger inom den naturliga tolkningens ramar. Finns där en dold aggression: Varför åtlyds inte mina direktiv?!

Betoningen kan inrymma betydelsen att personen ifråga hittills har livnärt sig på skräpmat, varför det är en veritabel sensation att vederbörande kommer hem med morötter.

Vi har för lite information för att veta säkert, men redan detta lilla utbyte mellan personerna omöjliggör genom sin betoning vissa tolkningar och gör andra högst troliga.

I ett specifikt historiskt läge kan också helt andra tolkningar framstå som givna. Kanske har nya rön från Lantbruks-

universitetet visat att morötter är giftiga. Eller så råkar de just nu vara symbolen för en främmande fientlig makt, varför de annars oskyldiga morötterna i har kommit att bära på laddade konnotationer. Detta kräver kulturell förståelse men framför allt förståelse av hur det känns att stå i relation till den främmande makten, och till folkopinionens bannbullor. Det kräver dialektisk förmåga – resonering, analyserande, vägande, tänkande, det som människan äger i rikt mått när hon är som bäst.

Andra betoningen:

Har *du* köpt morötter?

Även här öppnas för ett antal betydelser medan man stänger för andra. Den exakta intentionen går inte att fastställa. Den mest uppenbara betydelsen är att den frågande finner en påse morötter på köksbänken och undrar: "Hur har de hamnat där?" "Varför ligger det morötter här?" "Vem har köpt dem?".

Antingen träder alltså en Sherlock Holmesk slutledning fram: "Om inte du, så någon annan, för jag har då inte köpt dem." Eller så anmäler sig en Sigmund Freudsk deduktion: "Du brukar varken handla eller äta morötter, så vad har hänt med dig?"

I påståendet går det även att ana en viss fientlighet, förutsatt att det rimmar med förhistorien. "Har *du* köpt morötter?" kan betyda: "Har du börjat hjälpa till med hushållsarbetet nu plötsligt, det var då inte en dag för tidigt."

Tredje betoningen:

Har du *köpt* morötter?

Vad har vi här? En förhärdat snabbare som talar till en kumpan som fått för sig att börja betala för varorna? Eller är det en grönsaksodlare som frågar sin make varför han har betalat för morötter när hela trädgårdslandet är fullt av dem?

Fjärde betoningen:

Har du köpt morötter?

Den betyder: "Har du gjort det jag bad dig om?" I utsagens betoning ligger att de talande är överens om två premisser, dels att det var du:et som skulle handla, dels att det var morötter som skulle köpas. På något annat sätt kan saken inte förstås. Tolkningarna är aldrig oändliga och den här innebörden är fastlåst i just den här betoningen. Att morötterna inte skulle hämtas i ett grönsaksland behöver inte sägas ut. Frågan gäller om uppdraget att handla morötter har utförts.

Betoningen kan också rymma en undertext om tillit. Kan den frågande lita på att den svarande verkligen gör det som förväntas av honom? Frågan kan däremot inte vara en variant av innebörden "Var har du lagt morötterna?", för om det är detta man undrar ställs frågan i imperfekt: "*Köpte* du morötter?".



Morötter är morötter, jo! Men det finns många sätt att tala om dem, som en artificiell intelligens ännu inte förstår. Foto: iStock/Hyrma.

VÅR EGENART

Detta är vår egenart, vårt mänskliga liv, där varje liten vardags-situation bär på en värld av tolkning och avkodning med avseende på mål, mening och förståelse. Mål, mening och förståelse är vad roboten saknar. Och aldrig någonsin kommer den att få tillgång till det eftersom robotens essens inte är människans.

För att söka mål, mening och förståelse måste man veta hur det *känns* att förstå. Och för att veta hur det känns att förstå behöver man vara gjord av människomaterial. Man behöver veta vad det är att leva och minnas hur det känns att inte förstå. Vidare behöver man vara på det klara med vad människan håller kärt, vad som är viktigt och oviktigt och varför så är fallet. En perfekt tolkning av en situation kräver därtill att man vet vad individen framför en håller kärt och finner viktigt.

I allt som sker söker människan begriplighet. Begriplighet uppstår av mening. Begriplighet och mening kan bara uppstå ur en jämförelse med en normativ struktur i världen. Det gör att den helt fördomsfria varelsen inte kan tänka. Tänkandet måste ske utifrån en struktur, som ytterst är ett antagande om vad en människa är och vill.

Det här kommer roboten att lära sig rabbla, men rabblande är ingenting värt för människan. Ty vad AI än lär sig att processa, kommer AI aldrig att veta hur det känns lära sig saker och samtidigt drabbas av en svindlande känsla inför sina begränsningar, sin okunnighet och oförmåga, att rodna och svettas, att oro sig för att inte räcka till, att frukta, vändas, minnas. AI kan inte förstå välvilja eller illvilja, bara känna igen dess språkliga uttryck. AI kommer aldrig att överväga att behärska sig men se sig själv ge efter för lustarna, aldrig fundera på att anstränga sig men vara för lat för att orka. Sådant utgör stoffet för människans mest drabbande vardagserfarenheter, och eftersom AI är utestängd från dem ska AI aldrig ägna sig åt något som har med humaniora och samhälle att göra.

Vidare: AI kommer aldrig att kunna glädjas åt att ha lyckats lägga band på sig och njuta sötman i att slippa ångra ord som aldrig uttalades. AI vet inte vad ruelle är, kommer aldrig att känna ånger över en obetänksam handling.

VAD ÄR EN MÄNNISKA?

Frågan om AI är en fråga om varat. "Vad är en människa?", "Hur är det att leva?", "Hur bör livet levas?" är frågor som vi har slutat ställa oss, eftersom vi helst håller oss till frågor som kan beläggas med empiri i vetenskapliga studier. Tron att de existentiella och ontologiska frågorna är besvarade, eller saknar relevanta svar, är vad som har banat väg för övertygelsen om att artificiell intelligens kan föra mänskligheten till ett högre plan, alternativt utrota vår art.

Roboten har säkert många goda användningsområden inom medicinsk vård. Jag missunnar den inte att få arbeta i sitt anletes svett där. AI borde också vara en bra minröjare. Kanske kan den få syn på intressanta samband i stora datamängder. Men som variant av människa, som samhällsvarelse, ska robotarna hållas kort. AI är ett långsamt verkande gift, ett monster som dödar den rest av tänkande och själsliv vi har lyckats bevara.

AI är som människan när människan är som sämst: stumma fakta utan omdöme, rädd att använda sin erfarenhet på ett förnuftigt sätt, en varelse utan metafysik men som på grund av sin oförmåga att abstrahera, bedöma och göra korrekta hierarkier, lever i en absurd abstraktion.

FÖTTER, INTE RÖTTER

Det finns ett politiskt slagord som lyder: Människan har fötter, inte rötter. Sanningen är att människan har materiella fötter och metafysiska rötter. Roboten har varken rötter eller fötter. Den är i verklig mening rotlös. Denna frånvaro gör den olämplig för allt det som vi människor sätter värde på i fråga om mänsklig samvaro.

När vi börjar tro att AI är en bättre version av oss själva – mindre diskriminerande, mindre känslöstyr, mer oväldig och rättvis – och av det skälet försöker ersätta oss själva med robotar, bör vi tänka på några saker: Den urskillningslösa människan är motsatsen till en idealisk människa. Den urskillningslösa människan är en idiot. Den icke-kännande människan är en livsfarlig psykopat. Roboten är inte rättvis, däremot oförmögen att fatta hur man ska tänka för att komma fram till vilken mänsklig rättvisa som är rimlig i en viss situation.

Sokrates återkommer då och då till sin *daimonion* som talar om för honom när han ska avstå från att göra saker. Denna inre kraft talar aldrig om för honom vad han ska göra, bara vad han inte ska göra, vad han ska iakttaga försiktighet inför. AI har ingen *daimonion*. Däremot kan AI lära sig att i en evighet pladdra på om Sokrates daimonion så att det låter kunnigt.

FÖRNUFT OCH FÖRDOM

Under lång tid har västerländsk filosofi och det västerländska intellektuella livet ägnat sig åt kritik av omdömeskraften. Hänvisningar till förståndet har mötts med misstro, ty förståndet har ansetts vidarebefordra och försvara konventioner, tidens godtyckliga åsikter, klassens maktutövning eller gruppens ”privilegier”. Förnuftet har avfärdats som fördomens säte och som ett medel för auktoritär påverkan. Tankandet, förståelsen, resonemanget har stått tillbaka för faktakoll, siffror, mätresultat.

Det är därför inte att undra på att vårt innevarande samhälle vill ersätta människan med artificiell intelligens. När människan steg för steg avstår från sin mänskliga intelligens och strävar efter att bli en människorobot, har hon banat väg för den konstgjorda robotens frejdiga framfart.

Precis som en myra är bäst på att vara myra är roboten bäst på att vara robot. Och människan är bäst på att vara människa, det vill säga *moralisk meningsskapare*.

Frågor

Charlotte Erlanson-Albertsson, medicinsk kemi: Lysande Lena!

Tack! Vi talar om att AI inte hör hemma i humaniora. Men den hör inte heller hemma i mycket av medicinen. Det vi gör idag är mest att i översiktsartiklar sammanfatta tidigare forskning. Man gör inget nytt för att det blir allt svårare att tänka ut experiment, utföra experiment (till enorma kostnader). Vi har avhumaniserat oss genom att vi inte själva vågar göra experiment. Vi låter bilder och apparater göra dem och då förlorar man någonting: erfarenheter, iakttagelser, upptäckter som inte planerats från början kommer inte in i en datoriserad föreställning. Denna fantasi som är viktig i all verklighet, som gör att vi kan associera väldigt fritt. Är det någon som bjudit motstånd. Vad tycker folk när du säger att AI inte behövs?

Lena Andersson: I min sfär möts man snarare av invektiv än av argument. Man får en diagnos, till exempel bakåtsträvare, konservativ, reaktionär, att man inte förstår att detta är framtiden. I humaniorasfären är man väl ändå lite skeptisk. Men jag har inte hört många direkta reaktioner på det jag har sagt.

Joakim Wörnberg: Du nämnde Sokrates. Jag tar oss tillbaka dit och får kanske utmana lite. Jag håller med dig helt och hållet om att AI inte är sett ur det mänskliga (perspektivet?) Men det är väl Platons tolkning att Sokrates ställer sig skeptisk till att Platon börjar skriva ner kunskap. Nietzsche var skeptisk mot skrivmaskinen. Men vi speglar ju vårt tänkande i våra verktyg. Jag vet att du och jag delar en kärlek till reservoarpennan ... Kommer AI bli ett verktyg som hjälper oss bli mänskligare? Eller har vi hela tiden blivit mindre mänskliga?

Lena Andersson: Mindre mänsklig blir man när man hela tiden sönderdelar allting. Vi avfärdar saker med att ”det där går inte att veta”. Vi måste vänta på en undersökning, en forskningsrapport. Det är då vi gör oss till robotar. Vi har banat väg för AI genom att göra oss själva till robotar. Och vi beundrar dem för att de är bättre robotar än vi. Vi ska inte vara robotar!

Christer Löfstedt: Strålande tycker jag också. Men jag anar att du vill provocera lite. Och som ”lagom-orienterad” människa vill jag varken hylla AI eller säga att inte har i humaniora att göra. Utmaningen är väl att hitta AI:s roll? Vore jag forskare i humaniora kunde jag nog hitta något användningsområde. Jag vill inte dissä AI till 100 procent. Vill du modifiera din slutkläm? Eller ska vi kasta ut barnet med badvattnet?

Lena Andersson: Det kommer inte att hända, det som jag säger, så det spelar nog ingen roll. Jag bara hävdar att AI inte passar här. Det finns säkert något med datamängder, som vi kan använda, men vi måste ändå sitta där och tolka. Jag tror att det finns ett stopp för AI därför att tingen – även AI – har sin natur. Vad är meningen är frågan vi ska ställa oss. Vi har ju gått ifrån mening i vår tid idag så jag förstår att AI passar i vår tid.

Julia Ravanis: Jag håller med dig på flera punkter, men jag undrar hur du ser på möjligheten att använda AI som ett verktyg i konsten. Vi har mycket litteratur om hur det är att vara människa omgärdad av teknik ... till exempel fantasyböcker om robotar. Inom musik finns en mänsklig artist – Molly Herndon – som gjort en cover på Dolly Partons ”Jolene”, som hon låter en AI sjunga. Det handlar ju om kvinnlig avundsjuka på en vackrare kärleksrival och när AI-rösten sjunger den laddas originallåten med en ny betydelse, maskinens avundsjuka på människan. Hon leker med AI och

kan på så sätt belysa något nytt. Man kan alltså utforska nya frågor med hjälp av AI. Och lekfullheten är en kärna i konsten. Att förstå vad en människa är. Därför tror jag att AI kan ha en betydelse.

Lena Andersson: Jag håller med om att lekfullhet är ett konstens kärnuppdrag. Men då är det fortfarande människan som är herre. Det är vi som tänker, det är för oss som det är meningsfullt. Roboten är vår tjänare. Vi ska fortsätta att vara herrar, tycker jag.

Johan Richter: Jag förstår det som att AI får vara kvar i mitt område, medicinen, men hålla sig borta från humaniora. Jag är bara lite nyfiken på om du har lika dogmatiska åsikter om andra användningsområden. Hur är det med teknik, ekonomi, politik, samhällsbyggnad? Har du en lista med två kolumner? Hur sorterar du upp sakerna?

Lena Andersson: Det kanske jag inte har något svar på. Jag tror att vi vill ha mänsklig kontakt. Alla vill ringa ett nummer och få ett svar och klara av problemet. Vill man nå ett företag vill man framför allt kunna ringa! Jag förstår inte att inget företag förstår det. Jag har förbluffats av när man beställer flygbiljetter av AI och den föreslår att man ska ha 21 timmar på sig till Zürich! Eller Bokus ger mig förslag att jag ska köpa min egen bok. Det är så mycket AI inte har förstått, hur ska jag då kunna lita på den?

Allt blir väldigt opersonligt när man outsourcar allt till en robot. Jag håller med Peter Gärdenfors, att vi försöker göra oss själva till robotar. Vi skapar abstrakta system där det inte går att resonera sig fram till någon rimlighet och man kan inte klaga någonstans. Det är vi som är problemet. Vad är det för mening? Vad ska vi ha det till? Vem är intresserad? Jag vill mana till försiktighet, särskilt samhällsbyggare.

Germund Hesslow: Bara en kommentar: Forskning går inte att bedriva om man inte vet vad man vill veta. Vad andra kan behöva veta. AI kanske kan skriva romaner? Men AI vill inte kommunicera något särskilt. Själva poängen med konst är ju att man vill åstadkomma något hos den som lyssnar. Det kanske en AI kan efterlikna, men det ligger väldigt långt framme. Tänk på det när det talas om forskning att interagera med yttervärlden. Den forskning jag varit med om är fylld av ständiga improvisationer, just för att man har mycket i bakhuvudet som man vill och önskar.

Ola Wendt: Tror du att AI kommer att förstärka skillnaden ytterligare mellan naturvetenskap och humaniora?

Lena Andersson: Det kan man väl tänka sig. Humaniora och samhällsvetenskap är tolkande vetenskaper och de är nya, tidigare inrangerade under filosofin. De kanske inte passar i det system som naturvetenskapen ställer upp? En del forskare vill komma runt det genom olika mätningar, men de ger inte många svar.



Peter Gärdenfors och ???



??? och Kristina Lång.



Germund Hesslow, Lena Andersson och ???



Torbjörn Frejd och ???

???, ??? och ???



??? och ???

??? och Johan Stenström.





iStock
Credit: HOROGE

AI – hot och möjligheter



**Sophia Zackrisson,
moderator**

Professor i diagnostisk
radiologi

Varmt välkomna tillbaka till eftermiddagens föredrag! Artificiell intelligens förändrar i snabb takt hur vi lever, arbetar, lär oss och interagerar med samhällets institutioner. Från att ha varit ett verktyg för beräkning och databehandling har AI utvecklats till system som kan analysera, producera och fatta beslut på sätt som tidigare uppfattades som unikt mänskliga.

Stryk?

Inför uppdraget att tala till er så tittade jag tillbaka på vilka plattformar vi haft genom tiderna för kunskapsöverföring. Med dagens symposium tar vi ju faktiskt ett steg tillbaka till den muntliga kommunikationen. Kanske är det en paradox att vi fortfarande använder den när vi träffas här för att tala om det tekniska och avancerade. Eller så är det ett tecken i tiden att det är precis det vi måste göra för att fortsätta att vara människor? Vilket mycket av förmiddagens föredrag handlade om.

Efter varje steg i utformningen av plattformarna, från den muntliga kommunikationen till den skriftliga, från boktryckarkonsten, läskunnighetens utbredning och – så klart – det senaste



Enda sättet att lära sig är att själv använda sig av AI. Inom medicinen och specifikt radiologin, så har AI kommit väldigt långt. Det är oftast en smal AI med uppgift att peka ut var en tumör sitter i ett bröst till exempel. Och detta är redan implementerat i bland annat Sverige, vilket vi får höra mer om snart. Är det då ett hot mot det mänskliga mötet i medicinen? Jag kan bara säga att hittills sorteras det som inte är värdeskapande i vården snabbt bort. I den kontexten märks det snabbt om det tillför kunskap eller bara utgör ”pladder”. Vi vet vad vill vi ha och vi vill inte ha information om en sjukdomsrisk som ligger decennier fram i tiden och vi kanske ändå inte har någon bot emot.

århundradets utveckling av radio, tv, media, internet inte minst, så har man talat om att det nya kan vara en fara, men det kan också vara en fantastisk möjlighet att utjämna ojämlikheter mellan folk i världen. Det kan göra oss mer välinformerade och sprida goda budskap. Men lika väl kan det användas i annat syfte.

Eftermiddagens sessioner kommer att ta upp olika perspektiv på AI i undervisningen, i försvar och krigföring och i mänskligt arbete. Jag har funderat på det faktum att vi som samlas här idag inte direkt är tränade i AI utan har lärt oss det vi kan på annat sätt och kan komplettera AI:n eller roboten med annat kunnande. Men vad händer med våra framtida generationer, våra kommande studenter, läkare, militärer, som varit beroende av att vara i symbios med AI för att ta sina beslut? Vi vet inte hur det kommer att se ut. Men redan nu är det tydligt att nivåskillnaden mellan lärare och studenter och läkare och patienter utjämnats tack vare AI. Man kan vara en välinformerad patient när man kommer till sjukhuset och vi läkare måste kunna bemöta den kunskapen.



iStock
Credit: sigurcamp

Generativ AI och högre utbildning



Thore Husfeldt

Professor i datalogi,
Köpenhamn

”Jag är till åtlöje”, sade språkläraren med gråtkvävd röst. Artificiell intelligens baserad på maskininlärda stora språkmodeller påverkar många områden – på gott och ont. Teknikens inverkan på framför allt *undervisning* blev uppenbar för alla bara några dagar efter att dialogroboten ChatGPT gjordes tillgänglig för varje skolelev i november 2022.

Sedan dess har jag talat om detta ämne ett halvt hundra gånger på workshops, föreläsningar och symposier för skolor, gymnasier, utbildningar, universitet, forskare och ministerier i olika länder. Det var vid en sådan workshop som en lärare sade: ”Jag är till åtlöje.” Hon kände sig pressad att delta i ett spel för galleriet, där hon förväntades låtsas bedöma elevers uppenbart maskinskapade skriftliga uppgifter.

Andra sorgsna röster präglas av *ämnessorg*: den vemodiga insikten att årtiondens metoder och ämnestraditioner är föråldrade över en natt, att innehållet i mitt älskade ämne är irrelevant och att några av människans vackraste färdigheter nu är förvisade till automatiseringens avförtrollade tomrum.

Utbildningsväsendet känner åter ett stort handlingsbehov som följd av en teknologisk förändring. Det är dessvärre svårt att hantera, både därför att digitalisering i allmänhet är svår och därför att utbildning är särskilt komplex.

DIGITALISERING

Digitalisering är ett processförändrande ingrepp. Det kan i vissa sammanhang gå riktigt bra och i andra riktigt illa. Det går illa med garanti när man digitaliserar ett område man inte förstår eller talar ärligt om.

Framväxten av just generativ artificiell intelligens automatiserar en väsentlig del av utbildningssystemets kärnfärdigheter: produktion av nyskriven text med höga krav på form och ytkvalitet, med fokus på syntes, analys, reflektion och kreativitet.

Dialogroboten ställer en helt ny typ av lärare till förfogande: en sokratisk handledare som uppvisar tålmod, vänlighet, entusiasm, varsamhet och ämneskunskap i en grad som få lärare av kött och blod kan prestera. Roboten möter dig utan att döma och anpassar sig till just din kunskapsnivå och kulturella kontext.

Men vänta! Var inte självständigheten själva meningen: att man skulle skriva själv? Att man skulle lära själv? Digitaliserar vi nu bort utbildningens egentliga syfte?

Dessvärre är saken än mer komplex. Generativ artificiell intelligens hotar inte bara att digitalisera bort det uttalade, eller det ”offentliga”, syftet med utbildning, utan också de outtalade

Det sägs att syftet med ett system är vad systemet gör – oavsett det syfte systemet påstås ha. Och denna ofta välmenande öarlighet är i sig inget problem, förrän vi försöker digitalisera systemet. Då måste vi bli precisa, annars digitaliserar vi fel och optimerar berättelsen i stället för syftet.

Undervisningsvärldens misslyckanden med stora öppna nätkurser (MOOC) är ett exempel. Nätundervisning visade sig ”bara” vara resursslöseri. I jämförelse ger generativ artificiell

intelligens möjlighet till långt mer ingripande digitalisering, som kan ha direkt katastrofala konsekvenser. I synnerhet om vi inte förstår vad syftet med utbildning egentligen är.

SYFTET MED UTBILDNING

Många samtal om undervisning börjar med välmenande entusiasm över lovande didaktiska grepp, fascinerande nya resultat eller insiktsfulla undervisningsteorier. Men innan seminariets två timmar eller kvällens flaska rödvin är slut hamnar vi ofta i insikten att ingen runt bordet är ense om vad syftet med utbildning egentligen är.

Kunskap: Skåltalarnas utbildningsbegrepp – det som forskare och lärare framhåller i högtidliga sammanhang – är upplysningens och bildningstraditionens. Syftet med utbildning är kunskapsproduktion och -förmedling i mycket självständiga (”ensamma och fria”) institutioner; särskilt i de germanska länderna präglas vi av det humboldtska bildningsidealet, där det sanna och goda har egenvärde, och där människan formas till ett fritt, kritiskt, självständigt och tänkande subjekt.

Humankapital: I mötet med avnämrapaneler, politiker och skattebetalare formuleras en annan hållning. Utbildning bidrar till samhället. Forskningens roll är att framtida kunskap som kan skapa behandlingar, medicin, produktion eller teknologi, så att samhället blir bättre eller rikare. Utbildningens syfte är att öka befolkningens humankapital (sjuksköterskor, hantverkare, ingenjörer) med sikte på samhällsnytta och arbetsmarknadens behov.

Nätverk: Vad vi oftast ser tillbaka på från våra studieår är inte föreläsningarna eller kurslitteraturen – utan de relationer till andra människor vi skapade under vägen. Här mötte vi våra vänner för livet, vårt livs kärlek och studiekamraten som tjugo år senare ordnar en fin projektanställning eller skriver ett viktigt rekommendationsbrev. Utbildningssystemet – oavsett om det är internatskola eller selektivt högstatusprogram – är i detta



perspektiv en kuraterad mötesplats, en social arena, ett ”meet market”. Det skapar exklusiva gemenskaper av personer från ett passande samhällsskikt, som tryggt kan investera tid, tjänster och romantik i varandra.

Upplevelse: Men kanske är utbildning bara en del av upplevelseindustrin? Man skall trivas under sina bästa år, uppleva något, smaka på olika intresseområden eller pröva skilda sociala roller. Det är de extramurala upplevelserna, studiemiljön, fredagsbaren, spexet och utbytesstudierna som är väsentliga; systemet mäter trivsel, kundnöjdhet och tillgänglighet. Utbildning blir ett självrealiseringsprojekt som ger tillfredsställelse här och nu (i stället för i framtiden), som gynnar den enskilde (i stället för samhället) och som skall vara maximalt inkluderande.

Politik: Slutligen är utbildningsdomänen både ett politiskt verktyg och en plattform. Det är ett verktyg för regional-, kriminal- och självklart arbetsmarknadspolitik. Syftet med Lunds universitet var att göra Skåne svenskt; syftet med långa dagar i grundskolan är att frigöra föräldrabetskraft; och så vidare. Dessutom är tankar om utbildning förbundna med människosyn i så hög grad att utbildning är en oemotståndlig scen för att signalera sina grundläggande värderingar. Hela samhällsteorier springer exempelvis ur idén om den oskrivna tavlan; många

ideologier har utfallslikhet som måttstock eller ideal; och jag kan i ett huj avgöra hur hänsynsfull eller empatisk du är genom att lyssna till din syn på utbildning. Därför är utbildningsberättelsen en väsentlig hållningsmarkör i den politiska arenan. Oavsett syftet, vad är då framgångskriteriet? Även här måste man förhålla sig till mycket olika ideal:

Förbättring: Under idealet om universell förbättring skall alla bli skickligare. Ett gott utbildningssystem möter de lärande där de är och ger nödvändigt stöd så att alla upplever en väsentlig kunskapsökning – oavsett den enskildes förutsättningar och utgångspunkt.

Mångfald: I självrealiseringsbegreppet skall utbildningssystemet hjälpa den enskilde att finna sitt ”sanna jag”, utveckla brokiga och särskilda intressen, hitta inre motivation och utveckla unika perspektiv och insikter. Olikhet och mångfald är målet.

Likhet: I socialiseringsperspektivet är utbildningssystemets uppgift att framta homogena individer som är socialiserade in i vissa beteende- och umgängesformer, och har omfamnat särskilda ideologier, professionella normer och politiska värden. Dessa individer är likformigt pålitliga, intelligenta, arbetsamma, samarbetsvilliga och följsamma. Och hur verkar utbildningssystemet då?

Lärande: Den lärande är ett oskrivet blad som tränas i kunskaper och färdigheter i en varsam, systematisk och gradvis process. De uppnådda färdigheterna kan vara just utbildningens ämnesinnehåll – till exempel i sjuksköterske-, hantverks- eller ingenjörsyrkena. Alternativt tränas de studerandes analytiska förmågor genom studiet av ett för samhällsekonomin i grunden ointressant stoffområde, men de utvecklade generalistkompetenserna överförs sedan av den färdigutbildade till en arbetsmarknadsbestämd kontext.

Bortfall: Selektionsförklaringen betraktar utbildning som en utsorteringsmekanism. De mindre lämpade gallras successivt genom examinationer och de studerandes egna

utslutningsmekanismer i grupparbete och projektförlopp. Den som är både inkompetent och okarismatisk har ingen chans. Processen är nödvändigtvis långvarig, så att individerna som genomgår hela systemet kan signalera en attraktiv blandning av intelligens, uthållighet, ämneskompetens, karisma, behag, samarbetsförmåga, status och sociala färdigheter.

Antagning: Ett specialfall av bortfallsperspektivet är att urvalet sker redan vid antagningen till mycket selektiva institutioner – i stället för under själva utbildningen. Utbildningens innehåll och verkningssätt blir därmed utan betydelse; urvalet av framstående humankapital har ägt rum före första studiedagen. Musikhögskolan och tandläkarutbildningen representerar ytterpunkter för dessa mekanismer: om du kommer in på Musikhögskolans violinlinje är du sannolikt redan en mycket skicklig musiker med många övningstimmar från barndomen, medan tandläkarhögskolan just *inte* väljer studenter som har borrat mycket i sina vänners tänder sedan åttaårsåldern. De flesta utbildningar verkar genom en blandning av dessa metoder.

ATT DIGITALISERA UNDERVISNING

Den grundläggande bristen på konsensus om utbildningssystemet är normalt inget problem. Vi har funnit ett stabilt jämviktsläge som tillgodoser de flesta perspektiven genom att tapetsera över de värsta motsägelserna.

Men när man beslutar sig för att digitalisera undervisning – särskilt om man vill leda en organisatorisk förändringsprocess – kan man bara använda ganska få av de $5 \times 3 \times 3 = 45$ olika insikterna om vad systemet *gör* utan att verka cynisk. På grund av sociala och politiska mekanismer i organisationer och politik måste vi alltså förvänta oss att implementeringen av digitaliseringsprocesser, inklusive användningen av generativ artificiell intelligens, utgår från de mer sympatiska, ”offentliga” berättelserna om utbildning.

En avgörande skillnad är om man uppfattar utbildning som en

transformation eller som en selektion. I det humboldtska idealet – men också i socialkonstruktivistiska teorier med utgångspunkt i *tabula rasa* – formas människan genom att lära något. I motsats härtill står signaleringsteorin, där processens funktion är att välja ut lämpliga personlighetsprofiler utan att individen nämnvärt förändras.

Det kan tänkas att det inte är antingen/eller, utan både/och: några relevanta färdigheter har den färdigutbildade nämligen faktiskt och mätbart tillägnat sig under utbildningen (t.ex. franska, förvaltningsrätt, svetsning, blodprovstagning). Därtill kommer givetvis egenskaper (intelligens, arbetsamhet, behaglighet) som är oföränderliga drag i personligheten, och för vilka utbildningssystemet snarast verkar som en framkallningsvätska.

Detta är i grunden en sympatisk hållning: utbildning är både en mekanism för lärande och för selektion. Vi kan inte lära studenter att bli intelligenta, men vi kan lära dem franska.

Den bittra insikten är nu att generativ artificiell intelligens automatiserar just de kognitiva färdigheter som vi *kan* undervisa. Den kunskap vi med tillförsikt har kunnat lära en människa kan vi nu också lära en maskin. Omvänt är de egenskaper till stringens eller insikt som vi år 2025 fortfarande efterlyser hos stora språkmodeller just de egenskaper som vi traditionellt heller inte har kunnat lära ut. Varken den genomsnittliga studenten eller språkmodellen kan författa en innehållsligt lysande essä, men båda kan nu lära sig att följa ytnormer om korrekt språk och oklanderlig struktur. Det lärbara värdet av utbildning har automatiserats.

ANALOGIER OCH JÄMFÖRELSER

Effektmultiplikator: Som verktyg tycks artificiell intelligens baserad på stora språkmodeller för närvarande öka värdet av redan produktiva människor väsentligt. Analogin är till begreppet ”force multiplier” inom militärstudier: teknologi som mångfaldigar effekten av befintliga, välfungerande resurser.



Att göra fiolljud

Tillgänglighet

Låg

Hög

Resultatkvalitet

Beror på
användaren

Genomgående
utmärkt

Signalvärde: Användaren

... kan spela fiol

... kan lära sig
något svårt

... har tillgång till
en grammofoon



Till exempel blir skickliga mjukvaruutvecklare väsentligt mer produktiva genom att överlåta rutinuppgifter till artificiell intelligens i stället för till juniora medarbetare. Denna observation gäller också teknologin som lärandeverktyg: för den vetgirige, nyfikne, disciplinerade och intelligente är dialogroboten en helt ny möjlighet att lära ännu mer ännu snabbare. Det är däremot mindre klart hur teknologin påverkar individer, som har svårt att ta ansvar för sitt eget lärande och för vilka utbildningssystemet redan i utgångsläget inte fungerar.

Skivspelare: Jag kan frambringa violintoner med hjälp av två helt olika verktyg: en violin eller en skivspelare. Skivspelaren frambringar väsentligt vackrare toner snabbare och billigare, men "att kunna spela violin" är en annan färdighet än "att producera violintoner". I denna analogi är artificiell intelligens skivspelaren; förmågan att hantera denna teknologi är oberoende av förmågan att spela violin. Om syftet är "att frambringa vackra violintoner" är skivspelaren den överlägsna teknologin. Därtill är den moraliskt överlägsen: den är universellt tillgänglig och uppnår utfallslighet genom att frikoppla användarens musikaliska färdigheter från resultatets kvalitet. I värsta fall



Springbock.

ödelägges motivationen för, lusten till och tillfredsställelsen av att *själv* lära sig spela violin av skivspelarens existens.

Trampolin: Springbocken (*Antidorcas marsupialis*), en sydafrikansk antilop, har för vana att plötsligt hoppa högt i luften med alla ben samlade för att därefter fortsätta betet. Evolutionsbiologins signalteori förklarar detta till synes underliga beteende som ett så kallat "ärligt signal". En frisk antilop kan nämligen hoppa utan större ansträngning; gamla och sjuka kan inte. Ett rovdjur, dolt för antilopen, kan snabbt avläsa denna signal och avstå från att investera energi i att jaga den friska antilopen. Den väsentliga insikten är att signalen är billig att sända och att avkoda; evolutionen har funnit en mekanism som båda arter drar nytta av. Men signalen ("jag kan hoppa högt") är inte identisk med färdigheten ("jag kan springa länge"). Bedömning i undervisningssammanhang måste ha samma egenskap: det skall vara svårt att förgylla en färdighet eller egenskap, och billigt och rättvisande att avläsa. Till exempel trettio sekunders samtal på franska, eller bladläsning av ett violinsolo. Artificiell intelligens – där en text av oklanderlig ytteffekt är lätt att producera, och vars rättvisa bedömning skulle

vara *mycket* tidskrävande – är i denna analogi en trampolin: alla springbockar kan nu hoppa lika högt.

Plagiat: Utbildningssystemet har traditionellt betraktat plagiat med förfäran och ansett det vara ett väsentligt brott mot akademiska normer. Men plagiat krävde åtminstone någon form av engagemang med materialet. Även om det endast var för att leta på rätt bibliotekshylla efter en lämplig text att plagiera. Dessutom var plagiat ganska lätta att upptäcka genom att hitta originaltexten. Den moderna användningen av automatiserad plagiatkontroll och de konsekvenser som följer av avslöjade plagieringsfall visar hur allvarligt utbildningssystemet har tagit denna företeelse.

De flesta argument mot plagiat är också argument mot automatgenererad text – vi förväntar oss att författaren inte smyckar sig med falska fjädrar, att vi läser författarens egna åsikter och att texten innehåller originella idéer. Tillitsbrott och bristande proveniens är bara en del av förklaringen: även om vi föreställer oss att plagiatorn utrustade sitt verk med en ärlig källangivelse i stil med ”kapitel 2 har jag skrivit av hos N.N.”, skulle det troligen ytterligare minska vår respekt och vårt intresse för texten. Automatgenererad text är ännu lättare att producera än plagiat och dessutom omöjlig att upptäcka, så vi måste räkna med att den kommer att stå för en mycket stor del av akademisk och professionell produktion. Värdet av sådan text är av samma natur som ”ärligt plagiat”, som vi traditionellt haft svårt att etablera akademiska normer för. Det verkar sannolikt att textproduktion kommer att förlora status samtidigt som texterna blir bättre.

Vänskapstjänst: Utbildningssystemet har länge accepterat att gymnasieuppsatser skrivs av elevens högutbildade föräldrar, universitetsessäer författas av den studerandes gode vän, och att grupparbeten ofta består av en ensam högpresterande individ som drar de andra över mållinjen. Tror man att ”syftet med utbildning är lärande” verkar detta märkligt och

smått pinsamt. Håller man sig till mer nyktra förklaringar av utbildningssystemets syfte är inget av detta underligt: systemets uppgift är att säkerställa att det kulturbärande samhällsskiktet kan bevara sin position; arbetsmarknaden är intresserad av individer med sociala färdigheter och välkuraterade nätverk, som kan förmå andra människor att utföra framstående arbete åt en; grupparbete utjämnar kompetensskillnader och skapar illusionen om utfallslighet och undervisningseffekt; och så vidare. Systemet gjorde alltså vad det tyst skulle. Den som kan ”ringa en vän” har åtminstone en vän. Men generativ artificiell intelligens utjämnar även denna signal: ”ring en vän” är demokratiserat. *Gott så*, tycker jag, men även den cyniska förklaringen för en osjälvständig hemuppgifts signalvärde är bortautomatiserad.

I undervisningssammanhang är syftet med skrivprocessen inte textproduktion. Syftet är i stället, ex. att ordna tankar, att signalera skrivefärdighet eller förståelse för textens ämnesområde. Inget av dessa syften tillgodoses av att jag utlokaliserar skrivprocessen till någon annan – vare sig till min kloka studiekamrat eller till en maskin.

Generativ artificiell intelligens har underminerat många discipliners traditionella bedömningsgrund – nämligen omfattande skriftliga uppgifter framtagna utan tillsyn – oavsett om syftet med bedömningen är relevant återkoppling i en läroprocess, nyttig justering av undervisningsnivå och -progression eller rättvis examination.

UTBILDNINGSSYSTEMET

Utbildningssektorn förnimmar ett stort handlingsbehov i reaktion på generativ artificiell intelligens på minst tre olika områden: lärande, innehåll och bedömning. Dessa områden och frågor är väsentliga men också mycket olika. Var bör vi som institutioner lägga vår uppmärksamhet?

	Lärande	Innehåll	Bedömning
Effekt av GAI	okänd och ojämn	okänd	katastrofal
Institutionell påverkan	låg	låg	låg
Förstått	nej	nej	ja
Status	hög	hög	låg

Lärande: Det råder ingen enighet om i vilken omfattning tillgången till digital lärandeteknologi kommer att förbättra eller försämrar lärande, eller om den bara ytterligare ökar skillnader i läranderesultat. Men utbildningssystemet har historiskt haft mycket liten framgång i att få människor att lära sig mer. Det förefaller osannolikt att pedagogiken kommer att förstå hur man ”använder artificiell intelligens på rätt sätt”, än mindre att institutioner skulle kunna implementera denna metod och rulla ut den genom utbildningssystemet med fortbildning, metodändringar, och så vidare.

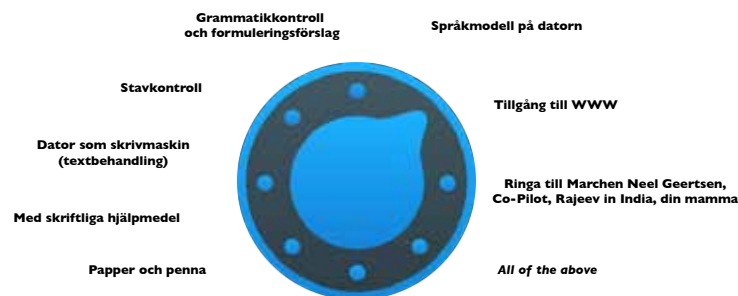
Min egen hållning är att hälsa AI (som alla andra informationskällor) maximalt välkommen och uppmana till ohämmad användning av denna teknologi *för lärande*, så länge vi håller den borta från *alla* former av bedömning. Detta innebär att vi måste förbereda oss på eventuella utfallsskillnader.

INNEHÅLL

En annan fråga är hur utbildningssystemet skall reagera på automatiseringen av färdigheter i symbolmanipulation, som traditionellt har varit en mänsklig domän och som särskilt universitet har valt ut, undervisat och certifierat: analys, sammanfattning, reflektion, översättning, skapande, o.s.v. Kommer hela discipliner att gå samma väg som typografutbildningen? Är ”att översätta Ovidius” eller ”implementera en databasmotor i Python” bortautomatiserat på samma sätt som ”att skörda med lie”?

Till skillnad från *lärande* kan institutioner påverka *innehåll* med precision och stor effekt, till exempel genom att lägga ned

Övervakad skriftlig tenta



[Tveksam till att ha med denna]

hela ämnesområden eller skapa nya. Innehållssidan är alltså institutionellt påverkbar. Men att fatta dessa beslut *korrekt* är svårt. Många akademiska discipliner har till exempel endast en svag koppling till den arbetsmarknad, där deras eftertraktade kandidater skapar värde. Vi kan göra något men vet inte vad.

BEDÖMNING

Frågan om giltig bedömning – en central uppgift för alla lärosäten – står i kontrast till de två andra temana. Effekten av generativ artificiell intelligens är enhetligt katastrofal för en stor familj av bedömningstraditioner. Hit hör självfallet längre individuella hemtentor, men också projektarbete i grupper och till och med skriftliga salstentor vid datorer som ger tillgång till en stor språkmodell via internet. Därtill är giltig bedömning ett *löst problem* i den meningen att vi vet *vad som måste göras*. (Nämligen: *färdighetsprov under uppsikt, inklusive muntliga förhör och skriftliga salstentor under kontrollerade former*.) Implementeringen är trista procedur- och förvaltningsfrågor om till exempel fysiska ramar (lokaler), autentisering och IT-styrning (tillsynsförda datorer). Allt detta är förvisso svårt men inom ramen för vad lärosäten just kan ta ansvar för.

Frågan om giltig bedömning kräver alltså vår omedelbara uppmärksamhet, även om den är grå och otrendig och lågstatus och inte harmonierar med vår självbild som vare sig Sokrates eller Humboldt. Vi bör lägga åtskillig organisatorisk uppmärksamhet på att implementera de uppenbara lösningarna för bedömning. Om vi inte ens mäktar denna förhållandevis beskedliga uppgift, förefaller det meningslöst att lägga energi på att diskutera aspekter där vi inte är i närheten av en klar problembeskrivning eller en lösning.

ANSVAR

Utbildningsväsendet *kan* lösa bedömningsproblemet, men det är en ledningsuppgift. Det är däremot inte en uppgift för den enskilde läraren. Att organisera workshops där enskilda lärare utan inflytande över de väsentliga processerna skall göras ansvariga för ”den goda hanteringen” av digitalisering är i bästa fall tidsödande, i värsta fall oansvarigt och hänsynslöst.

Slutligen: ansvaret för giltig bedömning kan aldrig läggas på eleven. Försätlig moralisering (”du fuskar väl inte?”) eller uppenbart oärliga uppmaningar (”du skadar bara dig själv”) är både verkningslösa och nedvärderande. Av etiska skäl bör vi aldrig inrätta system som primärt drabbar ärliga människor. Vi uppnår heller ingen trovärdighet genom att gallra bort de elever som faktiskt litade på oss. Som konsekvens riskerar utbildningsväsendet att förlora den sista resten av samhällelig välvilja och relevans för arbetsmarknaden genom att belöna oärlighet.

Ansvaret för att säkerställa giltiga bedömningsprocesser vilar av både moraliska och strategiska skäl på institutionen – inte på läraren och sannerligen inte på eleven.

Annars är vi till åtlöje.

Frågor

Nils Otto Sjöberg: Tror du att AI kommer att bli ett nytt ämne i skolan?

Thore Husfeldt: Jag tycker att datavetenskap behövs som skolämne. Många länder har det men de skandinaviska länderna ligger långt akterut. Men AI är bara en del av datavetenskap och det är datavetenskap vi ska lära oss.

Sven Åberg: Du säger att hemtentor kan man inte ha längre. Var går då gränsen: Projektarbeten, hemprojekt, examensarbeten? Det är ju stora delar av undervisningen som sker utanför skolsalen, mastersarbeten som innebär massor av poäng.

Thore Husfeldt: Ja, det är mycket svårt. För inläring är projektarbete alldeles utmärkt och jag har en viss sympati för problembaserat lärande. Grupparbete gynnar de starkaste eleverna i gruppen väldigt mycket, medan de svagaste lämnas efter. Jag tycker det är omoraliskt att göra grupparbete om man ser till vänstersidan av normalfördelningen.

Sven Åberg: Ja men om man följer upp efter examinationen med närmare frågor och ber dem resonera om svaren?

Thore Husfeldt: Det är stor skillnad mellan att lösa ett problem och förstå en lösning när man väl sett den. Tentamen ska vara individuell och giltig. Vi ska bli mycket bättre på att skilja mellan vår undervisningsroll och vår bedömarroll. Det senare har låg status och känns otrevligt.



Florence Nightingale

iStock™
Credit: TonyBaggett

Läkekonst bland intelligenta algoritmer



Kristina Lång

Docent i radiologi,
Lunds universitet

Läkekonsten skulle kunna beskrivas som konsten att på basen av vetenskaplig grund, beprövad erfarenhet och med ett empatiskt förhållningssätt, förebygga, diagnosticera och behandla eller lindra sjukdom. Samtidigt står sjukvården inför en AI-revolution, som utmanar just dessa egenskaper. Bill Gates har förutspått att AI kommer att ersätta en stor del av läkarnas arbete inom tio år. Frågan är alltså inte om utan hur AI kommer att påverka läkekonsten framöver.

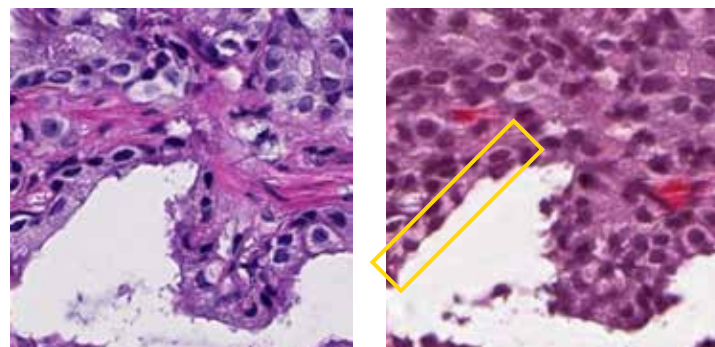
Sedan millennieskiftet har antalet medicinska publikationer om AI ökat 36-faldigt. De flesta berör ”smal AI” där algoritmer är tränade för en specifik uppgift, såsom taligenkänning eller bildanalys. Inom vården används AI redan som administrativt stöd. Taligenkänning gör det möjligt att diktera journalanteckningar direkt, vilket kan effektivisera läkarens arbete. Även om tekniken har funnits länge, fungerar den inte lika bra för alla. Precisionen i taligenkänning är ofta sämre för läkare med uttalade dialekter eller som inte har svenska som modersmål. Den digitala hjälpen

är alltså inte neutral utan bär på samma bias som den data den har tränats på.

ÖVERFLÖDIGA LÄKARE?

Inom bilddiagnostiken har AI haft ett stort genomslag. Redan 2016 förutspådde pionjären och sedermera nobelpristagaren Geoffrey Hinton att radiologer snart skulle bli överflödiga. Nio år senare vet vi att han hade både rätt och fel. AI kan i dag identifiera tumörer, frakturer och blödningar med imponerande precision – men har ännu inte ersatt radiologer. I stället har den snarare visat sig kunna utgöra ett stöd för röntgenläkaren i att effektivisera arbetet och för att minska missar. I dag finns över tusen godkända medicintekniska produkter som bygger på AI, varav majoriteten inom radiologi. De kan markera misstänkta fynd, prioritera akuta fall och sortera bilder efter risknivå. Ett konkret exempel är den svenska MASAI-studien, världens första stora randomiserade prövning av AI-stödd mammografiscreening. Studien omfattade över 100 000 kvinnor i Skåne och visade 29 % fler upptäckta cancerfall utan ökning av falskt positiva fynd, samtidigt som granskningsvolymen för radiologerna minskade med 44 %. Det är ett bra exempel på hur den smala AI:n kan användas för att både effektivisera vården och samtidigt leda till patientnytta.

På senare tid har generativ AI fått stor uppmärksamhet. Till skillnad från de smala AI-modellerna kan den skapa nytt innehåll som text, bilder och kod, som liknar det som människor producerar. Stora språkmodeller som ChatGPT och Gemini har visat förbluffande bra resultat och har redan "klarat" den amerikanska läkarexamen (USMLE). Språkmodellerna sägs nu klara USMLE bättre än människor. Är då ChatGPT att föredra framför en läkar-konsultation? Även om språkmodeller har tränats på en massiv mängd litteratur och kan svara adekvat på väldefinierade medicinska frågor, är den kliniska verkligheten ofta mer stökig. När språkmodeller har testats i realistiska patientfall, där symtom och labbresultat presenteras stegvis



I analysbilden till höger "hallucinerade" AI in celler som inte fanns i den ursprungliga röntgenbilden till vänster. Källa: Skånes universitetssjukhus.

och där handläggningen kräver kliniskt omdöme, minskar precisionen och läkarens resultat överträffar den generativa AI:ns. AI kan tyckas resonera logiskt, men den saknar förståelse för kontext. Den kan ge svar som verkar rimliga men som kan vara direkt felaktiga eller påhittade (hallucinationer). Utan läkarens expertkunskap är det svårt att avgöra när den har fel.

EMPATI

En central del av läkekonsten är empati och kommunikation, där vi människor intuitivt borde vara bättre än en maskin. Det verkar dock som detta påstående kan utmanas. I en uppmärksam studie publicerad i JAMA Internal Medicine från 2023 jämfördes hur empatiska svar patienter fick på sina frågor på nätforumet Reddit, besvarade av läkare respektive ChatGPT. De AI-genererade svaren bedömdes som både mer empatiska och informativa. I en annan randomiserad studie fick patienter med depression, ångest eller ätstörningar fyra veckors samtal med en AI-terapeut tränad på verkliga KBT-sessioner. Resultaten visade tydlig symtomlindring jämfört kontrollgruppen, som stod i kö till mänsklig terapisesession. Detta är lovande resultat, där AI kan



öka tillgängligheten till vården, när resurser är begränsade, men de får ändå tolkas i det begränsade sammanhang som studierna utfördes i. Den empatiska delen av vården är mer än så och det mänskliga mötet kan oftast inte ersättas.

Trots de lovande resultaten är AI fortfarande en teknologi under utveckling. De kliniska prövningarna är ännu få och de flesta randomiserade studier små. För att förstå teknikens verkliga värde krävs fler välgjorda studier. Än så länge kan AI inte ersätta läkekonsten, men den kan stärka den. Den kan effektivisera administrativa uppgifter, analysera och sortera datamängder och frigöra tid till det mänskliga i läkekonsten: patientmötet.

Frågor

Bengt Sjölund: Jag vet delvis av personlig erfarenhet att 20 procent av mammografierna visar falskt negativa resultat. Man hittar alltså inte tumörer som finns där. Nu har du med en fantastiskt stor studie, på 106 000 individer, visat att man kan hitta en massa tumörer vi annars skulle missat. Vilken effekt har det haft på sjukvårdspolitiken?

Kristina Lång: Vi befinner oss nu vid ändringar i policy och riktlinjer. Det har börjat implementeras av oss på fältet även om det ännu inte finns några klara officiella riktlinjer. Så är det i stora delar av landet, dock inte ännu i Skåne, men det är på väg, kanske redan om en månad? Det kommer snart både i Skåne och internationellt.

Arne Jönsson: Bill Gates behöver väl inte en försvarare, men det var en väldigt intressant intervju i New York Times. Han ska avsluta sin stora stiftelse inom 20 år. Han är ju mycket positiv till AI, mycket för att nå ut till de delar av världen som idag inte har några alternativ alls?

Kristina Lång: Det som han nämner om fri intelligens är viktigt. Det är en del av demokratiseringsprocessen att man kan förstärka låg- och medelinkomstländer. Det finns väldigt många olika applikationer inom det här området. Jag håller helt med Bill Gates om att det är en spännande utveckling.

Per Lundberg: Jag kan ingenting om vare sig läkekonst eller AI men tror mig förstå att ett mer holistiskt synsätt på människan och botandet av sjukdom diskuteras nu. Vi har ju varit väldigt ingenjörsinriktade men börjar nu mer förstå människan som

en helhet. Hur kommer den ansatsen att "gifta sig" med den nya AI-tekniken?

Kristina Lång: Jag håller inte med om att hela människan saknats i tidigare läkarutbildning och sjukvård. Vad AI-verktygen kan göra är ju att minska den administrativa bördan så man får mer tid för patienten.

Torbjörn Frejd, kemist: Finns det någon ansats med AI när det gäller trauma, vid snabba ryck som till exempel bilolyckor?

Kristina Lång: Ja, inom vilken del som helst inom medicinen finns det nya ansatser med AI: Automatisk tolkning av lab-parametrar, EKG, tester som görs i ambulanser, chatrobotar som kommit mitt i ett katastroftillstånd. Resultatet påverkas av dialekt, språklig brytning och emotionell påverkan.

Charlotte Erlanson-Albertsson: Fotograferar ni radiologer själva, eftersom du pratar om arbetslöshet?



Torbjörn Frejd

Kristina Lång: Röntgensköterskor/skötare tar bilderna men en följd av AI-analyserna är 44 procent mindre granskning för oss.

Eric Warrant: Jag tror att du redan delvis svarat på denna fråga. Men har ni provat AI-teknik på andra sorters tumörer?

Kristina Lång: Ja, det finns AI-modeller för upptäckt av mer eller mindre allt. Det pågår många studier för att förbättra träffsäkerheten och sortera patienter.

Sophia Zachrisson: Om ni bryter benet här i Lund på natten, så är det en AI som granskar bilderna först, som stöd för ortopederna som på morgonen kontrollerar om bedömningen var rätt eller fel.



iStock™

Credit: Apostolos Giontzis

Springer vi ifrån oss själva?



Peter Bovet Emanuel

Överstelöjtnant, doktor
Försvarshögskolan

Mänskligt beslutsfattande har alltid varit kärnan i militär verksamhet. I varje situation där tid, osäkerhet och risk kolliderar tvingas ledaren väga, tolka och agera. I dag förändras dessa förutsättningar dramatiskt. Digitalisering, sensornätverk och datadrivna system har skapat en informationsmiljö som växer snabbare än människan hinner förstå. I detta landskap framträder artificiell intelligens inte längre bara som ett hjälpmedel, utan som en potentiell partner i beslutsprocessen.

Den militära miljön fungerar som ett laboratorium för kognition under extrem press. Det som uppstår i krig, kris och snabbt föränderliga operationer blir i längden också relevant för civila beslutsmiljöer – för politisk ledning, krishantering och strategiskt viktiga organisationer. Det är därför inte bara en fråga för försvaret utan en fråga om mänsklig förmåga i stort: hur vi fattar beslut när tempot är högre än vår tankeförmåga, och hur vi bevarar vår förmåga att förstå sammanhang, bibehåller ansvar och säkerställer moraliska värden när tekniken börjar

fatta beslut åt oss i ett allt högre tempo.

I modern krigföring handlar beslutsfattande inte längre om en enskild strid eller frontlinje, utan om nätverk av samtidiga händelser. Mark, sjö, luft, cyber och rymd bildar tillsammans en domänöverskridande väv av system, sensorer och kommunikationer. Beslutsfattande sker inte längre sekventiellt utan parallellt och samtidigt, där varje beslut får omedelbara konsekvenser i flera domäner samtidigt.

Sensorer och vapensystem har tillförts AI – ibland för automation, ibland för autonomi. I båda fallen skapas effektivare informationsprocesser och snabbare beslutscykler. Den klassiska kedjan ”sensor-to-shooter” har blivit en ”sensor-to-decision”-kedja, där det inte längre är avståndet till målet som är den avgörande faktorn utan avståndet till beslutet.

Men med detta följer en paradox: mer data betyder inte automatiskt mer förståelse. Den mänskliga hjärnan är inte byggd för att parallellt bearbeta tusentals samtidiga informationskällor. Därför växer behovet av syntetisk kognition fram – inte som ett tekniskt modeord – utan som ett praktiskt svar på en operativ realitet.

I militära ledningssystem används redan AI för att filtrera, prioritera och analysera dataflöden. Men dessa system löser sällan problemet med att skapa mening. Mening uppstår i förståelsen av *varför* något sker – inte bara *vad* som sker. Här har människan fortfarande ett övertag, men också ett ansvar. För om vi låter tempot och volymen av data tvinga fram automatisering eller autonomi i fler led i beslutsprocessen, så riskerar vi att förlora förmågan till reflektion, kontext och ansvarstagande.

KOGNITIV FÖRSTÄRKARE

AI har länge beskrivits som ett verktyg, men i praktiken har den börjat fungera som en aktör i beslutsprocessen. I militära sammanhang används AI för allt från sensorfusion och logistikoptimering till taktisk planering och simulering av en



Bildtext. Foto: Försvarsmakten.

motståndares beteende. Systemen kan identifiera mönster som undgår mänskliga analytiker och föreslå alternativa strategier i realtid.

Nästa steg är utvecklingen av syntetisk kognition – system som inte bara analyserar data utan börjar resonera kring syfte, mål och kontext. Ett kontextmedvetet AI-system kommer inte bara kunna reagera utan också reflektera över förändringar i sin operativa miljö – och anpassa sig till de mest relevanta. I ett pågående forskningssamarbete försöker vi bygga just sådana system: arkitekturer som kan hantera både robotar och människor, optimera uppdrag i realtid – och kanske till och med anpassa sina modeller när verkligheten bryter mot förväntningar – eller självgenererade hypoteser som prövas med en hög frekvens.

Skillnaden mellan automation, autonomi och syntetisk kognition kan liknas vid skillnaden mellan ett verktyg, en aktör och en kollega. Automation utför, autonomi agerar, syntetisk kognition försöker förstå. Den sistnämnda kräver något som närmar sig ett slags digitalt omdöme – inte medvetande, men en förmåga att väga mål mot medel under förändring.

Det ställer flera nya krav, bland annat på hur vi definierar kontroll. I militära doktriner talas ofta om *uppdragstaktik* – att ge

underenheter frihet att själva styra över hur de löser sitt uppdrag inom en given och överordnad intention. Den principen kan bli vägledande även för hur vi utformar framtida AI. Det handlar inte om att ge maskiner fria händer, utan om att skapa system som förstår mänskligt syfte tillräckligt väl för att agera i linje med intentionen även när kontexten förändras. Det kommer också bli vitalt att utöver väl formulerade intentioner, målsättningar och värderelaterade principer även säkerställa funktioner för monitorering av deras efterlevnad.

DEN MÄNSKLIGA PARADOXEN

I militära miljöer är beslutsfattande ofta en fråga om liv och död. Därför har intuition, erfarenhet och så kallad militär fingertoppskänsla haft en särställning. Det är den tysta kunskapen – den som bygger på upprepade mönster, känsla för timing och förmåga att läsa situationer bortom det mätbara.

Men intuition är också långsam och sårbar. Den påverkas av trötthet, stress och förväntningar. Den mänskliga hjärnan är evolutionärt optimerad för överlevnad, inte för simultan analys av sensorflöden. AI-system däremot vilar aldrig. De är fria från känslor, minnen och bias – åtminstone i den meningen att de inte blir rädda, stressade eller trötta. Men just där ligger också deras begränsning. För i en värld där känslor, empati och värderingar styr de mest avgörande besluten kan kylig rationalitet bli farlig.

”Fog of war” – krigets dimma – symboliserar gränsen mellan vad vi vet och vad vi tror att vi vet. Den mänskliga förmågan att agera under osäkerhet bygger ofta på ofullständig information, men ändå med känsla för riktning. Ett AI-system kan ge oss klarhet och precision, men riskerar samtidigt att ersätta intuition med beräknad sannolikhet. Frågan är om den klarheten gör besluten bättre – eller bara snabbare.

När AI-systemen blir bättre än oss på att analysera – men fortfarande saknar förståelse för mänskligt värde – uppstår ett tomrum. Den syntetiska rationaliteten kan vinna slaget om

tiden, men inte kampen om meningen. Människan förblir därför oundgänglig – inte trots sina brister utan på grund av dem.

DATADRIVNA BESLUT

Allt fler beslut i moderna militära miljöer fattas på underlag som genererats, filtrerats eller simulerats av AI. Det kan gälla målidentifiering, riskbedömning, planering eller prioritering av resurser. Fördelen är tydlig: snabbare och mer informerade beslut. Men med det följer också en risk – att människan inte längre förstår grunderna för besluten hon fattar.

När beslutsunderlagen produceras i system vars inre logik är okänd uppstår det som ofta kallas black box-problemet. Vi vet att systemet fungerar – men inte varför. I en ledningscentral för ledning av militära operationer kan det innebära att en eller flera algoritmer föreslår en handlingsväg som tycks rationell men som bygger på, för människan, dolda antaganden eller felaktiga prioriteringar. Om människan då saknar insyn i hur slutsatsen uppstått blir hennes godkännande inte längre en akt av förståelse, utan av tillit. Tillit är i sig nödvändig, men när den blir blind förvandlas den till ett oönskat beroende. Människan reduceras då mot att bli ett av flera *moment* i ett automatiserat system, snarare än den som leder det. Snabbhet ersätter omdöme.

När beslutsprocesser blir för datadrivna finns risken att vi skapar system som fattar korrekta beslut i en felaktig värld. Den verkliga faran ligger kanske inte i första hand i att AI gör fel – utan i att den gör rätt enligt en logik vi inte längre delar eller riktigt begriper.

MÄNSKLIGT ANSVAR

Om syntetisk kognition är ett framtida bidrag till beslutsprocesser måste vi forma dess användning med våra värderingar. AI bör här inte efterlikna människan utan komplettera henne. Det kräver en ny ledningsfilosofi där samverkan mellan människa och maskin blir central, men där den mänskliga rösten fortfarande



Bildtext. Foto: Försvarsmakten.

är sista instans. Ju högre upp i det hierarkiska trädet av ansvar och påverkan – desto viktigare blir den nya ledningsfilosofins implementering.

Inom militär ledning talas det ibland om *command in control* snarare än *command and control* – en påminnelse om att ledning handlar om ansvar, inte bara styrning. I framtidens operationsmiljö behöver denna princip sannolikt utvidgas till *command in cognition*: en mänsklig ledning som inte bara övervakar, utan också förstår hur besluten växer fram, vad de baseras på och vilka konsekvenser de får.

Det kräver system som kan förklara sina egna slutsatser – så kallad *förklarbar* AI – och organisationer som tränar sina ledare och stabsmedlemmar i teknologisk förståelse. Det kräver också utbildning i reflekterande beslutsfattande, där teknikkompetens inte handlar om att kunna programmera, utan om att kunna tolka maskinens begränsningar och potentiella bias.

Ett AI-system kan prioritera effektivitet, men det kan (ännu) inte känna skuld. Det kan väga sannolikheter men (ännu) inte uppfatta rättvisa. Därför måste varje steg mot syntetisk kognition balanseras av ett steg mot ökad mänsklig medvetenhet och

professionellt omdöme. AI kan göra oss snabbare – men bara vi kan avgöra när det är rätt att sakta ner. Förutsatt att vi förstår. Förståelse kräver kunskaper, träning och prövning.

Det innebär också att utbildningen av framtida officerare och beslutsfattare måste förändras. Förutom taktik och strategi måste de tränas i teknologisk etik, kritiskt tänkande och kognitiv resiliens. Att leda med AI innebär inte nödvändigtvis att lita på systemet utan att förstå när det inte ska lita på. Därför är det också troligt att en ökad mängd anläggningar för simulationer kommer krävas för att möta ett växande behov av ny ledningsträning.

GEMENSAM KOGNITION

Vi står nu i ett skede där den militära kontexten inte längre är isolerad från samhällsutvecklingen – den är dess spegel. I en värld där kriser, konflikter och komplexa problem kräver snabba beslut blir frågan densamma överallt: hur förenar vi snabbhet med eftertanke, data med mening, teknik med ansvar? Framtidens beslutsfattande kommer sannolikt att vara en gemensam kognition – där människa och maskin delar analys, perspektiv och tempo. Det handlar inte om att välja mellan människa och AI utan om att skapa en syntes där maskinen hanterar det komplexa och människan definierar det värdefulla.

Men det kräver en medvetenhet om att teknikens logik inte är neutral. AI lär sig av det vi ger den – våra data, våra beslut, våra värderingar. Om vi inte vet vad vi lär den, vet vi heller inte vad den ger (lär) oss tillbaka. Därför måste utvecklingen av syntetisk kognition alltid vara förankrad i mänskliga normer, kultur och etik.

Filosofen Bernard Stiegler beskrev teknik som ett *pharmakon* – både botemedel och gift. Det gäller i högsta grad AI. Den kan stärka vår förmåga till förståelse, men också försvaga vår vilja till reflektion. Det är upp till oss vilken roll den får spela.

Att springa ifrån sig själv är inte ett tekniskt misstag utan ett mänskligt. Det sker när vi slutar reflektera över vad vi egentligen

vill uppnå med tekniken, när vi ersätter mening med effektivitet och ansvar med algoritm. Våra hjärnor behövs inte mindre i en värld av AI – de behövs mer medvetet. Syntetisk kognition kan bli en förlängning av vårt tänkande, men bara om vi låter den spegla det bästa i oss, inte ersätta det. AI kan hjälpa oss tänka snabbare, men bara vi bör avgöra *vad som är värt att tänka på*.



Frågor

Bengt Sjölund: Ska vi tolka dig så att svenska försvar makten är på väg in i en situation där "decision to kill" ligger hos AI och inte hos människan? Det har ju varit en stor debatt om det i den här beslutsloopen ska finnas en människa eller inte.

Peter Bouvet Emanuel: När jag svarar nu gör jag det som forskare vid försvarshögskolan och inte som anställd officer. Frågan kopplar till det man kallar meningsfull mänsklig kontroll. Det finns tre klassiska roller där: i loopen, på eller utanför loopen. Det är viktigt att vi balanserar "the decision to kill" för det finns redan system som gör det. Vid vissa system, som till exempel i vårt luftförsvar, så hinner vi inte vara med i den loopen. Men då talar vi om hot som annars kan utplåna den divisionsenhet som står vid batteriet. Luftvärnssystem ombord på större fartyg är ett annat exempel som kan överlåtas åt AI. När inkommande hot från robotar övergår mänsklig förmåga kan vi bara slå över till "automatic mood". Men allt handlar inte om att döda. Det fattas beslut om logistik och kommunikation till exempel. Vilka kriterier sätter vi upp för när en människa måste fatta besluten? Det är inte definierat ännu.

Ola Wendt: Om man tittar i krigshistorien så ser man att ju längre från dödandet som beslutsfattandet kommer så tycks det leda till en något större hänsynslöshet. Finns det då en diskussion om den hänsynslöshet som skulle komma ut ur en motståndare och hur man ska bemöta det?

Peter Bouvet Emanuel: Det klassiska exemplet är en drönpilot som sitter i ett land och slår till i grannlandet på mycket långt håll. Det blir allvarliga komplikationer, inte bara att etiken

förskjuts. De inblandade mår dåligt när de kanske alltför sent förstår vad de har bedrivit. Sverige och NATO får inte vara naiva. Människans natur är emellanåt mycket dyster och mörk, men vi får inte låta bli att beforska sådant som skulle kunna ske.



Julia Ravanis

Doktor i teknikhistoria,
Chalmers

Om man följer med i nyhetsflödet kring artificiell intelligens och jobbmärknaden är det lätt att bli deprimerad, eller i alla fall dyster. Återkommande inslag i medierapporteringen kring detta ämne är listor på vilka jobb som kommer att försvinna först (revisor, kommunikatör, programmerare), procentandelar av arbetsmärknaden som kommer att påverkas av AI (70 procent, 80 procent, och så vidare, beroende på hur man definierar "påverkan") och spekulationer kring vad alla dessa potentiellt arbetslösa, AI-ersatta, människor ska göra istället för det de gör nu. Budskapet är tydligt: AI tar våra jobb.

Det senaste decenniets expansion av så kallat smarta tekniker, "internet of things", och framsteg inom fält som maskin-inläring, som ligger till grund för mycket av det vi kallar artificiell intelligens, benämns ibland som den fjärde industriella revolutionen. Termen myntades vid en industrimässa i Hannover i Tyskland 2011 och har sedan dess ivrigt upprepats framförallt av frontfigurer inom tech-branschen. Symptomatiskt för de

tidigare industriella revolutionerna har varit just nya teknikers omdaning av arbetsmarknaden. Tittar man närmare på historien om teknik och arbete så dyker det upp paralleller inte bara mellan revolutionerande tekniker – utan också mellan andra karaktäristiska aspekter av de industriella revolutionerna; en storslagen men auktoritär retorik och ett onödigt passiviserande av den mänskliga arbetskraften.

DEN FÖRSTA REVOLUTIONEN

Den första industriella revolutionen inträffade i Europa under 1800-talet och är knuten till ångmaskinen. Den nya ångkraften möjliggjorde både storskalig produktion och transporter av varor längs de nya järnvägarna. Fabriker växte fram och ersatte alltmer den tidigare arbetsplatsen (i hemmen) för dem som producerade tyger, hantverksprodukter och andra varor. Det arbete som en spinnerska eller väverska tidigare gjort för hand hemma hos sig skulle nu utföras med hjälp av en maskin i en fabrik. Och maskinen gjorde det mycket fortare. Uppfinnaren James Hargreave, som uppfann spinnmaskinen Spinning Jenny 1764, fick sitt hus stormat av arga hantverksarbetare, som anklagade maskinen för att stjäla deras sätt att försörja sig.

Den första industrialiseringen gick hand i hand med kapitalismen och urbaniseringen: kapital ansamlades hos dem som ägde fabriker, medan fabriksarbetarna exploaterades, och människor från landsbygden ansamlades på platserna runt fabriker, det vill säga i städerna. Många av de gamla hantverkarna fick arbete, det var bara det att det var en annan slags arbete än det de tidigare utfört: mindre självständigt, mindre fritt, och mer mekaniskt och maskinellt.

Den andra industriella revolutionen följde på den första. När städerna växte och fabriker spred sig ökade behovet av att förse industrier och hem med gas, vatten, avlopp, kommunikationsmöjligheter och inte minst energi. För i slutet av 1800-talet fanns det en annan energikälla som höll på att



konkurrera ut ångkraften; elektriciteten. När städer och industrier elektrifierades innebar det återigen stora effektiviseringar. Nu behövde ingen längre skyffla in kol i ångmaskinerna, utan de kunde drivas med hjälp av elektricitet och därmed placeras ut på många fler ställen.

ROBOTAR SLAVAR ÅT OSS

Det arbete människan tidigare hade utfört med sin kropp sades nu kunna utföras av elektriska maskiner, och återigen spreds rädslan för att försörjningsmöjligheterna skulle förloras till robotar. Ordet ”robot” slog igenom efter att ha använts i en pjäs av tjecken Karel Čapek 1920, som handlar om robotliknande slavarbetare. Čapek satte ord på ett fenomen som då var högaktuellt: inte egentligen att maskiner blir som människor, vilket vi nu förknippar med ordet robot, utan snarare att de nya teknikerna tvingade fabriksarbetarna att bete sig som maskiner. Det löpande bandet är emblematiskt för den här epoken och förknippat med Henry Fords bilproduktion i början av 1900-talet (det fanns löpande band-teknik även före Ford, men de var inte

alls lika vanliga). Arbetare som monterade delar vid ett löpande band var tvungna att utföra helt monotona rörelser.

Detta nya sätt att arbeta på krävde en ny, mer komplex organisation – precis som det nya sättet att bo på, hoptryckta i städer, krävde komplexa samhällsstrukturer som avloppssystem och gasledningar. Ingenjören Frederick Taylor tillhandahöll en modell för en sådan organisation av arbete på verkstadsgolvet såväl som i kontorsmiljö. Taylors läror byggde på en strikt rationalitet, som reducerade arbetare till komponenter i produktionskedjan.

DATORER FÖR INTELLEKTET

De två första industriella revolutionerna sprang ur tekniker som förändrade kroppsligt arbete, och den tredje skulle göra detsamma med intellektuellt arbete. När den nya elektroniken möjliggjorde utvecklingen av de första datorerna, var det människans hjärna som sades kunna ersättas.

Ordet dator myntades i Sverige först 1969, av datalogen Börje Langefors. Innan dess kallades de nya maskinerna omväxlande för matematikmaskiner och elektronhjärnor. ”Matematikmaskin” var en väldigt rättfram benämning; de första digitala datorerna var rätt och slätt beräkningsmaskiner, som utvecklades under det andra världskriget för att lösa numeriska problem av militär karaktär, till exempel att rikta ballistiska missiler mot anfallsmål eller knäcka kodade underrättelsemeddelanden. ”Elektronhjärna” var ett namn som kopplade samman denna aldrig förut skådade beräkningsförmåga med mänsklig intelligens.

Datorpionjärer som Alan Turing började tala om artificiell intelligens redan på 1950-talet, när det fortfarande var en pärs för de tidiga programmerarna att få en dator att lösa en helt vanlig ekvation. En dator behärskade då ingen annan del av det som vanligtvis brukar ingå i definitionen av mänsklig intelligens än den logiskt-matematiska. (Det är ett omdebatterat fält, men



En IBM-avdelning i Kanada 1963. Foto: ???

en traditionell indelning är språklig, rumslig, logisk-matematisk, musikalisk, kroppslig, emotionell och social intelligens.)

Datorer blev snart en stor kommersiell affär, som snabbt anammades av verksamheter som inte varit lika berörda av ångkraften eller elektrifieringen som verkstadsindustrin: statlig administration, försäkringsbolag, banker. Kontorsarbetarna fick nu se sina jobb förändras drastiskt. Istället för att registrera och bearbeta information manuellt, var deras uppgift nu i allt större utsträckning att mata in och ut data i en ny maskin. Att köpa en dator för att strömlinjeforma administrativ verksamhet marknadsfördes på 1950- och 1960-talen som en revolutionär förändring, men tittar man på ekonomin för organisationerna som införde datorer i sina verksamheter syns ingen sådan. Teknikhistorikern Thomas Haigh, som har studerat just detta i artikeln *The Chromium-Plated Tabulator* från 2001, kallar datoriseringen av administrativa verksamheter under den här tiden för ”an ongoing and constantly increasing black hole in the budget”.



Eva Lindsten i arbete på FOA 1961. Foto: ???

SKAPADE NYA UPPGIFTER

Att mata in och ut data ur de tidiga datorerna var nämligen inget enkelt arbete. Det krävdes ett helt kompani av stanserskor, kontorsbud, maskinoperatörer och programmerare för att få en dator att göra det den skulle. Och bara hålkortsförbrukningen (all data lagrades på hålkort) motsvarade ofta en hel årslön.

Datorer har ju under 1900-talets andra hälft utvecklats otroligt snabbt. De har blivit mindre, snabbare, och allt eftersom erövrat fler områden av den mänskliga intelligensen. Trots detta är det svårt att inte tänka på misslyckade digitaliseringsprojekt som Skolplattformen (en digital plattform för lärare, elever och föräldrar som infördes i Stockholms kommuns skolor 2018) eller Millenium (ett gemensamt journalsystem för Västra Götalandsregionens sjukhus som infördes 2024) när man läser Thomas Haighs liknelse mellan datorisering och ett svart hål i budgeten.

Skattepengar läcker ständigt ut i sprickorna, eller i vissa fall kratrarna, som öppnar sig mellan digitala system och verkliga arbetsmiljöer. Precis som under den tidiga datoriseringen, finns det under den pågående AI-boomen en tendens att förbise allt det mänskliga arbete som trots allt fortfarande krävs för att hålla datorerna igång när man beställer nya tekniska system och lägger om arbetsmiljö och administration.

NYA OMRÅDEN EXPLOATERAS

Det rent materiella och manuella arbetet kopplat till AI och smart teknik finns också kvar, men det är mycket mer dolt idag än vad det var i 1960-talets datacentraler. Det är till exempel ofta geografiskt skilt från både programmerare och användare. Mineralerna och metallerna som krävs för att producera halvledare, transistorer och andra elektroniska komponenter bryts i gruvor i bland annat Kongo, i regel under omänskliga arbetsförhållanden. Dataanotering, det vill säga bearbetning och kategorisering av träningsdata till stora språkinlärningsmodeller och andra AI-tjänster, utlokaliseras till låginkomstländer där kollektivavtal och andra arbetsrättsliga skydd saknas. Många vittnar om digitala arbetsplattformar som är ogenomträngliga och alienerande och arbetsuppgifter som i värsta fall är traumatiserande (som att dagarna i ända gå igenom bilder och videor på barnporr och övervåld i censureringssyfte) och i bästa fall intellektuellt dränerande (ett ständigt flöde av bilder där man ska markera om det är en trappa eller en katt som visas upp). För att utföra den typen av arbete så snabbt och effektivt som möjligt blir det nödvändigt att anamma en maskinell logik – för att få lön för sitt arbete måste man till exempel svara på frågor på ett sätt som algoritmen bedömer som ”rätt” och därför bör man lista ut hur algoritmen vill att man ska tänka.

Exemplet ovan visar hur jobb inte bara försvinner i sviterna av en industriell revolution utan också skapas. Men vad är det för slags jobb som skapas? Om de två första industriella



Arbete med karaktärisering av inscannade bilder för AI-program. Foto: Andrew Norman Wilson.

revolutionerna manade människors kroppar att bete sig som maskiner, manar de två nästkommande människors intellekt att göra detsamma.

VAD VILL VI?

När man säger att "AI tar våra jobb" tillskriver man en agens till själva systemen, som döljer företagen och företagsledarna bakom systemen och som passiviserar alla dem vars jobb det handlar om. På vilka sätt arbetsmarknaden omdanas av ny teknik är inte bara en fråga om vad tekniken kan göra, utan också om vad vi vill att den ska göra. Vilken teknik behövs i ett samhälle? Vilka yrken vill människor jobba med och vilka yrken vill vi överlåta åt maskiner? Detta är politiska frågor. Men det är också frågor som kan översättas till institutionella kontexter, som inte nödvändigtvis måste besvaras på en nationell eller global nivå.

Vilken teknik tillåter vi i ett klassrum? Vad vill vi att studenterna ska lära sig och vilka moment i lärandet skulle det vara bättre att ta hjälp av AI för att göra?

Detta är ju ändå den grundläggande idén bakom all teknik-utveckling, hur revolutionerande den än må vara: att tekniken ska hjälpa oss bygga våra samhällen och leva våra liv.

Frågor

Ola Wendt: Blir det inte ett planekonomiskt samhälle om vi styr över om folk ska få skaffa sig en ny Iphone eller inte.

Julia Ravanis: Man pratar ofta från politiskt håll om utvecklingen som en autonom kraft som inte alls går att styra. Alla borde lägga sig i teknikutvecklingen genom att säga vad de vill ha.

Per Lundberg: Är det inte något kvalitativt annorlunda med AI-revolutionen jämfört med de tidigare?

Julia Ravanis: Den kommer att ombilda idén om vilka vi är och hur samhället ser ut. Jag påpekar för min del att det också behövs en kontinuitet.

Olle Söderman: Nu sker utvecklingen inom de stora tech-jättarna. Är det rimligt?

Julia Ravanis: Nej, det är en kraft vi bara kan svara på, inte styra. Techjättarna vill saker. Göra vinster, skapa nya innovationer, ibland även kringgå regelverk som begränsar dem, (göra oss beroende av dem, sa hon *inte*) Det behövs en motpol, en agens från politiskt håll. Det behövs tankar om hur vi vill organisera vårt samhälle.

Victor Öwall: Jag vill ifrågasätta lite det du sa på slutet om att människor jobbar med det de vill göra. Vi är kanske en privilegierad skara som får jobba med det vi vill. I Silicon Valley talar man nu om AI-arbetslöshet. Får vi fler arbetslösa? Eller hur förändras arbetsmarknaden?

Julia Ravanis: Det är naturligtvis jättesvårt att sia om. Som

historiker ser man ju på hur det varit förr. Ekonomin beror ju inte bara på den tekniska utvecklingen. Nya sektorer där Ai introduceras drabbas nog av arbetslöshet. Men på lång sikt tro jag inte att utvecklingen på ett deterministiskt sätt är på väg mot AI-arbetslöshet.



Några medverkande och planeringsansvariga för symposiet (?).

Kungl. Fysiografiska Sällskapet i Lund

Stortorget 6, 222 23 Lund

Tel: 046-13 25 28

E-post: kansli@fysiografen.se

www.fysiografen.se

© Kungl. Fysiografiska Sällskapet i Lund och författarna

Grafisk formgivning: Lönegård & Co

Omslagsbild: *Tänkaren* av Auguste Rodin. Foto: Metropolitan Museum of Art.

Foton från symposiet: Kennet Ruona

Tryck: Balto Print, 2026

ISBN: 978-91-985039-9-9