



Jan Lindvall

## Artificiell intelligens – människa och maskin

### I korthet

För att stärka människans roll i den digitala miljön krävs att vi satsar på att utveckla digital kompetens. ■

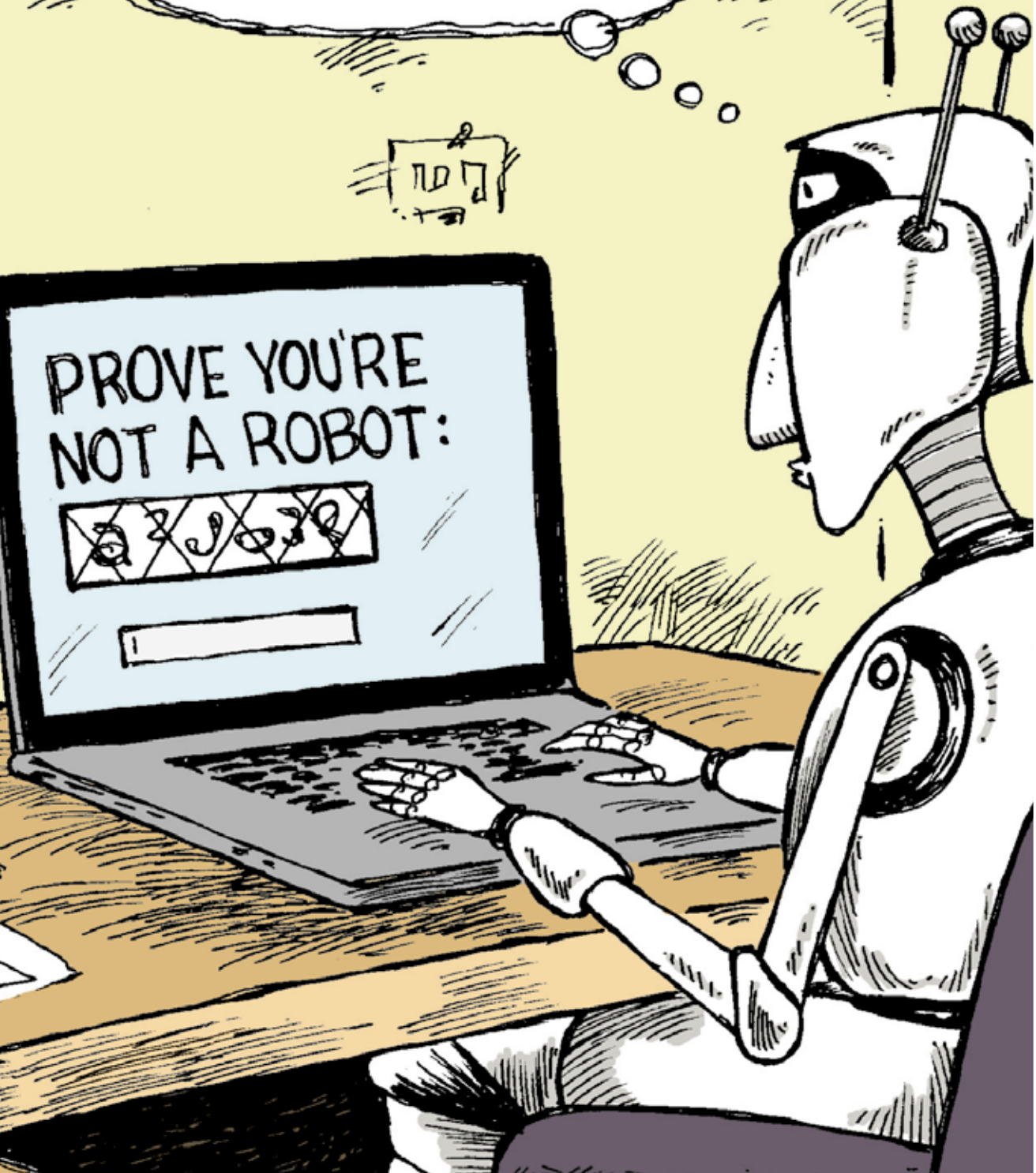
**Relationen mellan människa och maskin** är av central betydelse för organiserandet av verksamheter. Därför förvånar det inte att introduktionen av ny teknik präglas av både förhoppningar och rädslor. I dag märks detta tydligt i diskussionerna om artificiell intelligens (AI). AI förväntas lösa många viktiga samhällsfrågor, exempelvis inom miljö och medicin. Samtidigt finns farhågor om att en okontrollerad utveckling av AI kommer att leda till att smarta maskiner snart överträffar människan i intelligens och förmåga. Maskinerna kommer att ta över, och människan blir maskinens slav. Så vad är denna maskinella intelligens, som lockar och skrämmer? Finns det verkligen läroaktiga – intelligenta – maskiner? För att få perspektiv på den frågan är det lärorikt att blicka tillbaka historiskt på de förväntningar och farhågor som förekommit tidigare kring datoriseringen och de "tänkande maskinerna".

Inledningsvis bör det påminnas om att historien visar att tekniska förväntningar ofta vida överträffar omfattningen av de praktiska tillämpningarna – vi har fortfarande inte självstyrande bilar i allmän trafik. Först kommer drömmen om teknikens alla möjligheter, långt därefter följer, eventuellt, ett mer omfattande användande och ofta då med modifierad funktionalitet. Detta leder till att vi riskerar att överskatta teknikens disruptiva kraft på kort sikt, men underskatta dess långsiktiga konsekvenser. Ett uttryck för att utvecklingen tar tid och att det inte alltid blir som tänkt är att det fortfarande, trots att det var redan 1956 som Herbert Simon tillsammans med andra introducerade begreppen *artificial intelligence* och *thinking machine*, inte finns någon generisk AI att tillgå. Det går fortfarande inte att föra över maskinellt lärande från ett tillämpningsområde till ett annat. Dagens AI-lösningar är fortfarande enbart utvecklade för specifika tillämpningar, exempelvis ansikts- och röstigenkänning. Utvecklingen tar alltså tid.

OMG, THAT IS SO OFFENSIVE!



PROVE YOU'RE NOT A ROBOT:  
 a  z  j  o  k  e





I en reflektion kring AI många år senare gav Kasparov IBM:s teknik ändå sitt erkännande: han såg den som intelligent”

Utvecklingen är inte heller alltid linjär. Dagens framgångsrika lösningar för maskinellt lärande baseras på tänkesätt och lösningar som delvis har varit bortglömda men som först nu kan tillämpas praktiskt eftersom datorernas bearbetningskraft är större och tillgången till stora datamängder är mer gynnsam. Gamla tankar kan alltså realiseras i och med att tekniken hinner i kapp föreställningsförmågan.

Teknikutvecklingen är alltså varken linjär eller autonom. För att teknik ska få genomslag krävs att de rätta sociala förutsättningarna finns på plats. Om vi vänder blicken mot just AI och maskinlärande måste vi till att börja med fundera över vad en teknik de facto förmår prestera ”intelligensmässigt”. Ett vanligt sätt att göra det är att pröva teknikens funktionalitet och kapacitet mot ett idealiserat mänskligt tänkande. Inom datorisering och AI är Turing-testet från 1950 det mest välkända. Där prövas en dators intelligens utifrån vilken förmåga ”maskinen” har att föra ett intelligent ”mänskligt” samtal. Om människan inte förstår att det är en maskin den pratar med ses det som tecknet på att maskinen uppvisar någon grad av mänsklig intelligens.

Under åren har ett flertal Turing-tester gjorts. En tidig framgång var Weizenbaums ”Eliza”, en dator som föreföll kunna föra enklare samtal. Maskinens synbart mänskliga beteende baserades på en förprogrammerad logik där vissa centrala ord i en fråga returnerades till frågeställaren. Det skapade igenkännande och skenbar förståelse hos frågeställaren. En lärdom Weizenbaum drog av projektet är att även välutbildade professionella användare lätt förleddes till en övertro på teknikens intellektuella förmåga.

I andra tester prövas teknikens funktionalitet mot sällskapsspel. Under lång tid var schack, med sina tydliga mål, väldefinierade regler och omfattande kombinationsmöjligheter, provostenen för datorers intelligens. Det väckte därför stor uppmärksamhet när IBM:s dator Deep Blue 1997 vann över världsmästaren Garry Kasparov. Bortförklaringarna var omedelbara: Kasparov hade spelat ovanligt dåligt och maskinen vann på överlägsen minnes- och bearbetningsförmåga. Vinsten var därför mer beroende av ”maskinell styrka” (stor datakraft) än skarpare intelligens. Snart återställdes ordningen. Kasparov vann påföljande matcher. I en reflektion kring AI många år senare gav Kasparov IBM:s teknik ändå sitt erkännande: han såg den som intelligent.

Senare, 2011, vann IBM:s dator Watson mot två stormästare i Jeopardy. Segern symboliserade ett viktigt utvecklingssteg. Nu lämnades eran av fristående datorer (*computational era*) för dagens sammankopplade, kognitiva era (*connectionism*). Maskinen Watsons ”tänkande” blev med det mer likt

## Om författaren

**Jan Lindvall** är docent vid Uppsala universitet och forskar om digitalisering och verksamhetsstyrning. Han undervisar på masternivå och i doktorandutbildning kring analys och organisatoriskt beslutsfattande. ■

dagens föreställningar om mänskligt tänkande. Neurala nätverk bestående av sammankopplade, processande, neuroner blev den vägledande metaforen. I maskinens neurala nätverk genomför de många distribuerade neuronerna var för sig enkla beräkningar som sammantaget resulterar i komplexa lösningar. Med det utvecklades Watson i riktning mot en ökad "förståelse" av mänskligt språk. Det öppnade även vägen för ett associativt "tänkande". Jeopardy bygger ju på att svaret är givet, det är frågan som ska formuleras och då gäller det att kunna associera.

Även efter Jeopardy-segern kom det snabba bortförklaringar. Än en gång sades det handla om överlägsen teknisk kapacitet, inte maskinell intelligens, och Watson vann främst på snabb reaktionstid – Watson hade chansen att besvara fler frågor än motståndarna. En annan kritik gällde att maskinen inte stod ensam mot mästarna. För att serva Watson inför och under tävlingen behövdes tillgång till ett mycket stort antal kompetenta mänskliga understödjare. En tredje kritik kom från filosofen John Searle. I sammanhanget är Searle mest känd för tankeexperimentet "The Chinese Room", som ifrågasatte Turing-testet. Searles tankeexperiment går i korthet ut på att vi utgår från en standardmodell för informationshantering: indata, bearbetning och utdata. Vid bearbetningen finns det regler och instruktioner att följa. Om en person som är okunnig i kinesiska nu sätts i rummet och "matas" med kinesiska ord kommer hen snart att kunna leverera ett meningsfullt svar i andra ändan. Enligt Searle betyder det inte att personen bör ses som intelligent: hen följer ju bara instruktioner och har inte lärt sig att *förstå* kinesiska. En semantisk syn handlar om att sammanhanget – kontexten – har en avgörande betydelse. För att förstå kinesiska måste vi förstå mening och användning av språket, inte bara följa översättningsregler. Information måste beaktas i ett relevant sammanhang för att tolkas väl. Searle hävdade att Watson saknade en sådan kontextuell förståelse. Den saknade dessutom intentioner (mål) och kunde därför inte betraktas som en tänkande, lärande, intelligent maskin.

Det senaste speltestet av stor betydelse är den match som Googles dotterbolag DeepMinds maskin AlphaGo gick 2016 mot den sydkoreanske mästaren Lee Se-dol i det kinesiska brädspellet Go. Go är ett strategispel som erbjuder omfattande handlingsmöjligheter. Med större spelplan (19×19, jämfört med 8×8) och enkla regler är kombinationsmöjligheterna mångfalt större än i schack. Valmöjligheterna sägs vara fler än det finns atomer i universum. Än en gång vann maskinen. Men denna gång restes inga invändningar mot segern. Snarare mottogs den som en indikation på att maskinen äntligen

närmade sig människan när det gäller förmåga till intelligens och lärande.

AlphaGos vinst innebar ett paradigmskifte i datorutvecklingen. Nu visade en multinivålösning för algoritmer och maskinlärande sin kraftfulla funktionalitet. I en sådan utformning kommer varje analytisk nivå att betraktas som en preliminär (sannolik) informationsmodell bestående av indata, bearbetning och utdata, där utdata från en lägre nivå blir en något mer förklarad indata på nästa, högre, nivå. Efter hand har dessa mekanismer mellan nivåer blivit alltmer oförklarliga för den mänskliga erfarenheten och förståelsen. Därigenom har det skett ett skifte i synen på hur lärande och intelligens utvecklas maskinellt. Med begreppet *deep learning* har en ny epok inom AI inletts. I modellerna utvecklas initialt preliminär kunskap genom att de tidiga modellerna/algoritmerna exponeras för mycket stora datamängder. Allt bearbetas och modifieras därefter av nya AI-algoritmer som utvecklas av datorn själv – algoritmer som gör att modellens använda sannolikheter kontinuerligt justeras utifrån gjorda erfarenheter ("feedback", "lärande") på underliggande nivåer. Detta betyder att algoritmerna i *deep learning* skriver sig själva och utvecklas utan att några mänskliga händer är inblandade.



Detta betyder att algoritmerna i *deep learning* skriver sig själva och utvecklas utan att några mänskliga händer är inblandade"

Innebär framgången för AlphaGo att maskinen nu har överträffat människan i intelligens och därmed på sikt gör människan kunskapsmässigt överflödigt? Troligtvis inte – det finns ju fortfarande ingen generell AI. Även om ambitionen med utvecklingen av deep learning är att reducera behovet av mänskliga insatser – och efterföljaren AlphaGo Zero har visat på stora framgångar här – går det inte att bortse från behovet av människor.

Detta för oss alltså tillbaka till det jag konstaterade ovan: tekniken och samhället utvecklas tillsammans. För att förstå att människan och specifika organisationers situation inte kan lämnas utanför AI-utvecklingen är det bara att se närmare på dagens två huvudformer av maskinlärande: övervakad respektive oövervakad inlärning (supervised/unsupervised machine learning).

Utmärkande för den allra vanligaste formen – övervakad inlärning – är att det finns ett uttryckligt mål gentemot vilket en modell tränas och sedan utvärderas. Här intar mänsklig expertis en mycket viktig roll, då människor inleder och övervakar lärprocessen. Det är experter inom den aktuella kunskapsdomänen som inledningsvis avgör vilken betydelse (sannolikhet) faktorerna i det använda datasetet ska viktas med. Det behövs alltså en inledande föreställning (teori) som efter hand prövas och därefter utvecklas genom deep learn-

ing, och mänsklig kompetens behövs även för att tolka algoritmens resultat.

Inom den oövervakade varianten, som fortfarande är en mindre del av maskinlärandet (men som AlphaGo Zero, AlphaGos efterträdare från 2017, utvecklades kring), sker lärandet utan behov av inledande mänsklig kunskap inom den domän där maskinlärandet sker. Det finns inte heller något fördefinierat slutmål för modellerna. Lärandet är helt maskinellt, en design som innebär ett mycket snabbare och mindre energikrävande lärande. Så lärde sig AlphaGo Zero "på egen hand", på fyra timmar, att spela schack på en sådan nivå att den vann överlägset mot marknadens bästa schackprogram. Detta betyder också att det inte behövs någon inledande teori att utvärdera prestationerna mot. I stället identifieras mönster utifrån stora datasetet där data inom grupperna är homogena men skillnaderna mellan grupperna är tydliga.

Det är ett maskinellt lärande i denna riktning, kopplat till förväntningar om ett framtida möjligt användande av mångdubbelt kraftfulla kvantdatorer, som kritikerna varnar för. Maskinen kommer då oundvikligen att vinna mot människan. Detta skulle kunna bli framtiden om kvantdatorer utvecklas enligt de högt ställda förväntningarna.

Även om det ibland kan förefalla som att människor i maskinella miljöer präglade av deep learning inte längre är viktiga, är det nog fel. Det finns fortfarande ett stort behov av kompetenta människor. Människor med väsentlig domänkunskap (kunniga inom ett specifikt fält) är avgörande för hur algoritmer utvecklas, väljs och tolkas. Människor kan ställa krav på och ifrågasätta kvaliteten i de stora dataset som är grunden för allt maskinellt lärande. Det är också människor som har intentioner och ställer de viktiga frågor utifrån vilken maskinen sedan utvecklar fördjupad kunskap.

För att ytterligare stärka människans roll i en alltmer digitaliserad miljö behövs därför kontinuerlig kompetensutveckling. Ett viktigt inslag bör vara utvecklingen av digital kompetens, en kompetens i övergripande frågor som dataarkitektur, datakvalitet och informationstolkning. Här har universiteten en viktig roll att fylla – en roll som ännu i hög grad tyvärr är lite förbisedd. ■



Det finns fortfarande ett stort behov av kompetenta människor”

---

#### Litteratur

- Boden, M.A. (2018), *Artificial intelligence: A very short introduction*. Oxford: Oxford University Press.
- Cantwell Smith, B. (2019), *The promise of artificial intelligence*. Cambridge, MA: MIT Press.
- Denning, P.J. & Lewis, T.G. (2020), *Technology adoption*. *Communications of The ACM*, 63(6), s. 27–29.
- Gärdenfors, P. (1996), *Fängslade information*. Stockholm: Natur & Kultur.
- Polson, N. & Scott, J. (2018), *AIO: Hur artificiell intelligens fungerar*. Göteborg: Daidalos.
- Sejnowski, T.J. (2018), *The deep learning revolution*. Cambridge, MA: MIT Press.