JAN CEDERVALL

AI-paradoxen

På senare tid har jag läst några artiklar och böcker som alla berör frågor kring artificiell intelligens, datorns teoretiska möjligheter och Gödels teorem. Det är Sten Lindströms och Ingar Brinks kapitel kallat "Artificiell Intelligens – Tankar utan innehåll" i antologin *Huvudinnehåll* utgiven av bokförlaget Nya Doxa, vidare boken *Språket, medvetandet och världen* av Hans Blomqvist som kommer på bokförlaget Novapress, Sten Henrikssons artikel "Det klarar datorn aldrig!" i *Forskning & Framsteg* nummer 4:1997, Jan Sandreds kolumn "Schackdatorer är meningslösa" i *Datateknik* nummer 20:1997 samt Marcus Bjärelands inlägg "Lite smartare maskiner mål för AI-forskningen" i *Datateknik* nummer 1:1998. Det har varit mycket läsvärt, inspirerande och tankeväckande, så inspirerande att jag själv finner mig manad att göra ett inlägg.

Jan Sandred menar att AI-forskningen är både dum och meningslös. Man får intrycket av att AI helst borde jämställas med astrologi. Jan Sandred skiljer mellan vad Lindström och Brink betecknar som teknologisk AI och en mer abstrakt kognitiv AI. I teknologisk AI-forskning försöker man efterlikna eller härma intelligent beteende genom att exempelvis utveckla vinnande schackspelande datorer. Målet för den här forskningen är inte djupare än illusionsteknik för illusionsteknikens egen skull. Jag håller helt med om att detta är både andefattigt och förhållandevis meningslöst.

I den abstrakta kognitiva AI-forskningen är det först när datorn agerar och skapar som ett resultat av egna medvetna tankar, egen vilja, egna känslor och egen intuition som man har lyckats och kan tala om intelligens. Här duger inga illusoriska tricks. Komponerar datorn musik måste det vara ett uttryck för egna tankar och känslor. Spelar den schack måste den kanske inte vinna men den måste veta om att

spelar och tycka det är kul. Denna mer abstrakta form av AI-forskning är enligt Sandred om möjligt än mer meningslös än teknologisk AI, ty den ligger utanför datorns teoretiska möjligheter. Jan Sandred påpekar att datorer arbetar enligt instruktioner i program som gör dem helt bundna av oföränderliga regler och åtminstone i teorin helt förutsägbara. Man får väl hålla med om att det är omöjligt att tillskriva en helt förutsägbar maskin egen fri vilja och intuition och vad skulle den ha tankar och känslor till när dessa ändå inte skulle kunna påverka det redan förutsagda beteendet. Nej man måste nog ge Sandred rätt i att drömmen om abstrakt AI utgör höjden av meningslöshet.

Över till något mer meningsfullt. Sten Henriksson påpekar att ett vanligt nybörjarfel som man stöter på i programeringsundervisning är program som aldrig stannar. Jag minns mina egna misstag som nybörjare då datorn fastnade i en programslinga där villkoret för att gå ur slingan aldrig uppfylldes. Vore det inte en lämplig övningsuppgift för lite mer erfarna dataelever att göra ett diagnosprogram som kontrollerar om ett program kommer att stanna och om det inte kommer stanna visar var i programkoden det kommer spåra ur. Vad händer nu om en busig elev byter ut diagnosutskrifterna? Den utskrift som säger att programmet kommer att stanna byts mot en evig slinga. Den utskrift som skulle tala om var programmet skulle spåra ur plockas bort. I stället stannar programmet utan någon utskrift. Vad blir resultatet om det modifierade program testas på en kopia av sig självt? Jo om det borde stanna så kommer det gå in i en evig slinga, men om det hamnar i en evig slinga, då borde det ju stanna! Vi har hamnat i en paradox, som utgör kärnan i en rad likartade bevis av Alonzo Church, Alan Turing med flera från senare delen av 30-talet och framåt, vilket tvingar oss att inse att vårt antagande att det gick att göra ett diagnosprogram som förutsåg hur ett godtyckligt program skulle bete sig, det antagandet var fel, ty i det allmänna fallet är datorer inte ens teoretiskt förutsägbara. Sten Henriksson redogör mer detaljerat för de uttryckliga delarna av resonemangen i ett av de centrala bevisen. Nej, övningsuppgiften var nog inte så lämplig, den var omöjlig medan den abstrakta AI:n kanske inte är så omöjlig som låtit påskinas; åtminstone är datorer inte teoretiskt förutsägbara som Sandred hävdade.[15]

15 En del kanske tycker sig ha fått bekräftat vad de länge misstänkt att datorer sällan gör vad man hade förväntat sig. Tyvärr måste jag göra den som

Förutom att datorer är teoretiskt oförutsägbara, så kan även teoretiskt förutsägbara program vara oförutsägbara av mer pragmatiska skäl. Vi kan jämföra med när vi lagar mat. Ibland spelar det inte så stor roll i vilken ordning vi gör saker och ting, om vi skär upp gurkan eller tomaterna först påverkar inte hur salladen smakar, medan det vid andra tillfällen har avgörande betydelse, om vi sätter på ugnen först efter att vi tagit ut pajen blir den nog inte så lyckad. Om vi är flera som ska laga mat tillsammans så kan det hända att skärbrädan är upptagen när gurkan ska skäras, men det går kanske lika bra att vispa grädden medan man väntar på att skärbrädan blir ledig. Genom att låta den ordning i vilken vi gör saker vara godtycklig så kan det gå fortare och smidigare än om vi har en bestämd ordning och tvingas vänta på varandra. Likadant är det i datorernas värd om vi avsiktligt låter någon faktor vara godtycklig till exempel den ordning i vilken olika beräkningar utförs så kan det gå smidigare och vi får resultaten snabbare men vi riskerar samtidigt att få in en viss oförutsägbarhet och resultatet kanske inte blir exakt det samma varje gång vi kör ett program. Små variationer i exempelvis datorns temperatur kan påverka tiden och därmed ordningen för beräkningsprocesserna på ett kaotiskt och i praktiken oförutsägbart sätt. När de faktorer som är godtyckliga är oförutsägbara så kan givetvis hela processen bli oförutsägbar.

Jan Sandred påpekar att mänskliga intelligenta resonemang inte bara är rationellt logiska utan även intuitiva och att intuitionen är kaotisk. Jag vet inte om Sandred vill att vi ska dra slutsatsen att datorer inte kan vara kaotiska, men vill han det är slutsatsen fel. Det är inte otänkbart att kaos spelar en viss roll i vår intuition men jag tror varken matematiskt analytiskt kaos eller för den delen filosofiskt kaos utgör huvudförklaringen till vår intuition. Intuitionen är en för oss oförklarlig känsla för vad som är rätt eller fel. Vi fattar beslut om en oviss framtid om vilken vi ej kan veta vad som är rätt eller fel ändå tycks vår intuition ofta leda oss rätt i våra val, vi känner intuitivt vad

---

eventuellt tycker sig ha fått vatten på sin kvarn lite besviken. Nästan alla vanliga datorprogram som man kan köpa eller hitta på nätet är förutsägbara, ett av få undantag är *Game of Life* som för en del spelöppningar blir oförutsägbart. Sedan är det tyvärr en annan historia att med få undantag kan programmen inte leva upp till av leverantörerna utställda förväntningar.

som är rätt eller fel, intuitionen vägleder oss i vår kreativitet, i kärlek, i moral och etikfrågor bland annat, områden där rationaliteten verkar otillräcklig. Jag tror att en huvudförklaring till vår intuition är förändringar i det algoritmiska dynamiska strukturstabilitetsmåttet hos vår viljefunktion. När vi är på väg att handla fel eller göra felbedömningar så sjunker strukturstabiliteten och vi får en intuitiv känsla av att något är fel, omvänt när algoritmisk dynamisk struktur och stabilitet verkar öka får vi en intuitiv känsla av att vi är på rätt väg. Om denna gissning rymmer någon riktighet vet jag naturligtvis inte, men den känns i vart fall något mer förklarande än kaoshypotesen och inte meningslös att vidareutforska.

Någon som AI-kritiker ofta hänvisar till är Kurt Gödel. Bakgrunden är följande; logiken och matematiken skördade oerhörda framgångar i början av 1900-talet och nya specialdicipliner som mängdlära, bevisteori, talteori och beräkningsteori föddes och utvecklades. Man började fråga sig om det fanns några gränser för de formella metodernas framgång och var gränserna i så fall gick. Att bevisa att en utsaga var sann eller falsk kunde vara nog så knepigt, men hur bevisade man ett ickeresultat, att en utsaga var obevisbar. Kurt Gödel var först med ett signifikant ickeresultat eller "negativt" resultat – en remarkabel bedrift, sedan dess har det kommit många negativa resultat.

Principen är följande, det är betydligt lättare att hamna i problem än att ta sig ur dem. Dessutom blir det bara värre ju påhittigare och djärvare man är, visserligen blir man bättre på att ta sig ur problemsituationer men man hamnar också i fler. Annorlunda uttryckt: ju kraftfullare ett formellt system är desto fler olösbara problem kommer att finnas inom systemet. Aha! triumferar AI-kritikerna. Det finns alltid vissa problem en dator inte kan lösa och aldrig kommer kunna lösa. Så sant, men vad AI-kritikerna tycks glömma är att dessa olösbara problem inte bara är olösbara för datorer, de är olösbara över huvud taget, även för oss människor.

Det finns även en något annorlunda form av kritik med koppling till Gödel. Lindström och Brink formulerar denna kritik tydligast. Det finns begrepp och resonemang som vi människor använder oss av i vårt mer intuitiva problemlösande som inte går att formalisera i en bevisligen konsistent ändlig axiomatisering. Det senaste är en viktig distinktion. Gödel med flera har inte sagt att dessa begrepp och resonemang är omöjliga att formalisera. Vad som sagts är att det är

omöjligt att formalisera om man samtidigt ska upprätthålla krav på till exempel konsistens, men vem tror att vi människor jämnt är konsistenta, sunda och rationella. Vad man inom AI vill är just att undvika fullständighet, konsistens och sundhet och i stället söka det oberäkneliga och irrationella utan att det för den skull blir vansinnigt eller absurt. Den traditionella datavetenskapen och klassiska logiken har helt enkelt andra mål än AI-forskningen, varför många av de mer traditionella kraven, begränsningarna och resultaten inte är tillämpbara eller relevanta.

Inget kan äga rum utan en föregående orsak, säger Jan Sandred. Huruvida detta är sant beror på vilken betydelse man lägger i de ord som ingår i utsagan, så för resonemangets skull har jag inga problem att gå med på det men jag är inte säker på att jag förstått poängen. Om slutsatsen är att detta leder till begränsningar för datorer till exempel att de inte kan vara spontana så måste väl den kritiken gälla även oss människor om inget kan undslippa sin föregående orsak. Om poängen i stället är någon slags allomfattande fatalism, så blir det svårt att ens för resonemangets skull att gå med på utsagan.

Det finns flera andra inriktningar på AI-forskning som ej tagits upp här. Marcus Bjäreland tar upp det faktum att de flesta AI-forskare inte sysslar med renodlade illusionstrick och det är bara en liten klick som hänger sig åt att spekulera kring abstrakta formaliseringar av djupa kognitiva processer, huvuddelen av forskningen och forskarna har en betydligt mera pragmatisk inriktning på att göra maskiner lite smartare. Det handlar om en föga spektakulär variant av tillämpad datavetenskap som gör datorerna lite mer användbara. Frågan om den abstrakta kognitiva AI:n är möjlig, handlar dock egentligen inte så mycket om datorer utan snarare om oss själva. Både vi och datorer är universella Turingmaskiner; frågan är om vi är någonting mer än Turingmaskiner. Något mer, som i grund och botten är oförklarligt och därför meningslöst att försöka förklara. Detta är en öppen fråga, men det är lite svårt att se hur ett svar skulle kunna se ut. Om vi lyckas skapa en dator som är medveten, känslig och intelligent, hur får vi i så fall reda på det, hur vet vi att någon annan än vi själva har ett medvetande? Denna fråga är det centrala temat i Hans Blomqvists bok *Språket, medvetandet och världen* och frågan sätts i boken in i ett mycket större, belysande filosofiskt sammanhang. Boken är mycket läsvärd och rekommenderas varmt men den löser inte problemet, den

ger inte svaret på frågan. Att Turingtestet inte är till mycket nytta är något som Jan Sandred riktigt konstaterar. Jag har tidigare framfört vad jag kallar synkronicitetstestet som ett alternativ. Det går ut på att försöka upprätthålla synkronitet med en oförutsägbar process utan möjlighet till synkronisering. Detta test ger en bättre indikation än Turingtestet men mer än en indikation får man inte, det besvarar inte frågan om datorn har ett medvetande. Om vi inte kan skapa medveten artificiell intelligens, så står vi förmodligen inför en ännu svårare eller i vart fall minst lika svår fråga. Hur visar man att medveten intelligens är oförklarlig och därför inte ens teoretiskt kan skapas med artificiella metoder, hur bevisar man att det finns något oförklarligt över huvud taget? Kan det vara så att frågorna om den abstrakta kognitiva AI:ns möjlighet respektive omöjlighet inte går att besvara? Men hur visar man i så fall det?

Är frågan om vi människor är något mer än maskiner dum? Vi människor eller vissa av oss i varje fall tycks ha fötts med okuvlig nyfikenhet, som driver oss till att ställa frågor och söka svar. Huruvida detta är meningsfullt är upp till var och en att bedöma, men för oss som har denna nyfikenhet går den inte att undvika hur dum den än må vara.