KUNGL. TEKNISKA HÖGSKOLAN - CSC

Kan en dator vara intelligent?

Christoffer Rydberg
Simon Janghede
2011-06-14

Handledare: Lars Kjelldahl
Examinator: Mads Dam

Abstract

Artificial Intelligence (AI) is commonly known today as something you would see in a Science Fiction movie with intelligent robots and machines. The things that are implemented today and people refer to as AI are widely used all around the market in places such as games, robots and web based computer scripts. The question is though if these scripts, programs and "smart solutions" can be called *intelligent*? Can a computer ever be *intelligent*? To be able to answer this question you would have to define what AI actually means. An approach to this has been done by several professors and doctors specialized in AI but certainly everyone has their own opinion.

In this paper we aim to present and conclude the arguments from the opposing sides of the battle between the sides of "Is artificial intelligence possible?", discuss these arguments and present our own ideas about it. Our aim is to present our ideas without forcing them to the reader. This meaning that we leave room for the reader to judge, keep his/her opinion or to have a sudden change of heart throughout the discussion.

Innehållsförteckning

1. Introduktion

1.1 Bakgrund

Många kanske förknippar Artificiell Intelligens (AI) med Science Fiction och tänker på en avlägsen framtid när robotar har tagit över världen, men faktum är att AI redan idag är ett välanvänt begrepp. De flesta som gillar datorspel har nog hört det användas när man talar om datorstyrda spelare och det är bara ett av många exempel. AI används brett för att beskriva system som är självlärande eller kan analysera situationer och utifrån det ta smarta beslut. Frågan är bara hur rätt det är att kalla dessa system intelligenta. Såhär beskriver nationalencyklopedin beskriver begreppet AI:

> **artificiell intelligens,** AI, dels intelligens som man tillskriver t.ex. ett datorsystem, dels ett forskningsområde inriktat mot att konstruera datorsystem som uppvisar intelligent beteende. Syftet är att på konstgjord väg efterlikna hjärnans förmåga att dra slutsatser, planera, lösa problem, förstå naturligt språk, lära etc.
>
> *Källa: Nationalencyklopedin. Hämtat 25:e mars - http://www.ne.se/artificiell-intelligens*

Utifrån ovanstående definition känns det naturligt att ifrågasätta till vilken grad intelligent beteende vittnar om faktisk intelligens. Vore det inte mer korrekt att säga att intelligensen ligger hos programmerarna, som talat om för programmet exakt hur det ska bete sig, snarare än hos programmet i sig? Var drar man då gränsen mellan att vara intelligent och att bete sig intelligent? Såhär ser det ut om man slår upp begreppet intelligens i nationalencyklopedin:

> **intelligens,** förstånd, begåvning, tankeförmåga, skarpsinne; förmåga till tänkande och analys.
>
> *Källa: Nationalencyklopedin. Hämtat 25:e mars - http://www.ne.se/intelligens*

Vi fastnar här för det första ordet: ”förstånd”. Hur kan man vara intelligent om man inte förstår vad man gör? Vi vill påstå att intelligens är något som uppstår tillsammans med förståelse och därmed en medvetenhet om den egna existensen. Det är under detta antagande om vad intelligens egentligen är som vi fortsätter i vår uppsats på jakt efter den intelligenta maskinen.

1.2 Problemformulering

Man brukar tala om svag respektive stark AI. Svag AI definieras av Ray Kurzweil som maskinell intelligens som motsvarar eller överstiger den mänskliga intelligensen för specifika uppgifter. Varje gång man skickar ett e-post ringer ett mobiltelefonsamtal är det smarta algoritmer som får rutt på informationen. AI-program kan diagnostisera hjärtsjukdomar, flyga och landa flygplan, eller göra automatiserade investeringsbeslut. Det finns oändligt med användningsområden.[1]

Stark AI är artificiell intelligens som motsvarar eller överstiger den mänskliga intelligensen inom alla uppgifter, där man ställer sig frågan om det är möjligt att skapa en maskin som kan lösa alla problem människor löser med hjälp av sin intelligens. Denna fråga är den som AI-forskare är mest intresserade av att svara på. Den definierar tillämpningsområdet för vad maskiner kommer att kunna göra i framtiden och styr AI-forskningens riktning. Räcker det med att maskinen beter sig som att den tänker, eller måste den verkligen kunna tänka? Såhär beskriver John Searle, en amerikansk filosof och forskare, begreppet stark AI:

> *[...]* the computer is not merely a tool in the study of the mind, rather the appropriately programmed computer really *is* a mind in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states.
>
> *Källa: Internet Encyclopedia of Philosophy. Hämtat 12e April 2011 - www.iep.utm.edu*

Datorer kan i dagsläget lättsamt utföra kraftfulla beräkningar som för en människa skulle ta mycket lång tid. Tvärtom finns det saker som för en människa är helt triviala men för en dator otroligt komplicerat att utföra. Beror dessa skillnader på att vi ännu inte har lyckats hitta algoritmer för de processer som möjliggör vårt intellekt? Eller är hjärnan helt enkelt för avancerad för att vara algoritmisk? Denna typ av frågor utgör grunden för vår uppsats som har som mål att ge klarhet i frågan: ”Kan en maskin vara intelligent?”.

1.3 Syfte

Syftet med uppsatsen är att ge en bild av hur långt man har kommit inom AI-forskning och diskutera hur långt man kan komma. Vi kommer beskriva några av de AI-system och projekt som finns i dagsläget samt jämföra den mänskliga hjärnans prestanda med dagens datorer. Slutligen kommer vi presentera de argument som finns för och emot att en dator kan vara intelligent, diskutera dessa argument och presentera våra egna idéer och slutsatser. Vårt mål är att presentera våra idéer utan att tvinga dem till läsaren. Det råder delade meningar om ämnet och det är ingen som med säkerhet vet hur framtiden kommer se ut.

DEL I

2 AI idag

Idag används termen *Artificiell Intelligens* inom en rad olika områden. Några exempel är robotik, datorspel och smarta scripts på hemsidor. Denna typ av AI är såkallad svag AI, som till skillnad från stark AI endast simulerar intelligens utan att egentligen skapa det. Dessa AI-system är algoritmiska och är med andra ord programmerade att följa ett visst mönster för att lösa olika uppgifter. Frågan om stark AI kan uppnås med algoritmiska lösningar diskuteras senare i uppsatsen.

> **algori´tm,** följd av instruktioner för ett beräkningsarbete som i ett ändligt antal steg löser ett beräkningsproblem och därmed kan utgöra grunden för ett datorprogram.
>
> *Källa: Nationalencyklopedin - http://www.ne.se/algoritm*

Tidigt gick AI-forskningen ut på att skapa algoritmer som imiterar steg-för-steg hur en människa resonerar när hon försöker lösa ett pussel eller att dra logiska slutsatser. Under slutet av 80- och 90-talet hade forskningen lett till framgångsrika metoder för att lösa problem där faktorer såsom *osäkerhet* och *ofullständig information* spelar in. Detta gjordes med hjälp av koncepten kring sannolikhetslära. Nedan beskrivs lite av den AI som finns idag.

2.1 Robotik

Som vi nämnt tidigare tolkar ofta människor smarta maskiner, lösningar och finurliga algoritmer som intelligenta. Idag finns smarta robotar som anses ha mycket avancerad AI. De är programmerade att lösa en viss uppgift på egen hand, gärna smidigt och snyggt och ofta genom att imitera mänskligt beteende och på så sätt verka intelligenta i betraktarens ögon. Stora utmaningar inom området robotik och AI är objektmanipulation och navigation. Dessa utmaningar har dessutom underliggande faktorer som att veta var man är, veta vad som är runt en och hitta en väg till sitt mål.

Ett system som kan tyckas intelligent är exempelvis IBM's Deep Blue[2] som år 1997 slog den dåvarande världsmästaren Garry Kasparov i Schack. Det var en sex omgångar lång match som slutade i två vinster till Deep Blue, en till Kasparov och tre matcher oavgjort. Under den andra matchen berättade Kasparov att han ibland såg djupt intellekt och kreativitet i maskinens pjäsrörelser och påstod att mänskliga schackspelare hade hjälpt maskinen under spelets gång – och därmed ”fuskat”. Det var första gången i historien en dator slog en regerande världsmästare.

Stanford University's tävlingsbidrag med vilken de vann DARPA[i] Grand Challenge med år 2005 är också ett exempel på en mycket avancerad och smart konstruerad robot.[3] Tävlingen går ut på att göra det bästa förarlösa fordonet och få det att köra en lång och svår off-road-bana på kortast tid. Att fordonet ska vara förarlöst innebär självklart att det inte får vara fjärr- eller radiostyrt. Fordonet ska alltså köra helt på egen hand.

Vinnaren av 2007 års DARPA.

Banan var 212 km lång och fordonet körde denna sträcka i off-road på 6 timmar och 53 minuter med en medel-hastighet på 30.7 km/h.[4] Vid nästa lopp två år senare var det helt annan terräng – det var storstadstrafik med allt vad det innebar.[5]

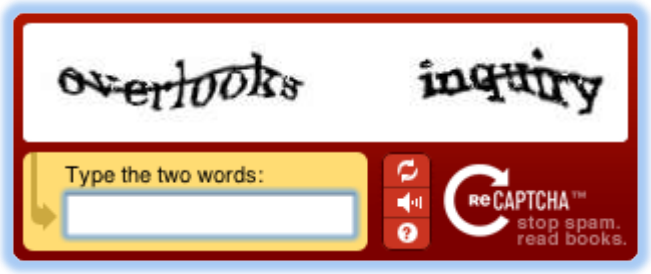
Det finns många robotar som är extremt imponerande men vi kan tyvärr inte berätta om alla. En till som vi gärna vill nämna är datorn som i Februari 2011 vann amerikanska Jeopardy mot de två Jeopardy-mästarna Ken Jennings och Brad Rutter. Datorn kallas "Watson"[6] och är konstruerad av IBM. Datorn är på papper bara en "frågelösningsmaskin" men är enligt utvecklarna mycket mer än så. När de utvecklade Watson så var de tvungna att tackla ett spel some inte bara behövde encyclopediskt minne, utan även förmågan att trassla upp hackiga och otydliga påståenden. Detta tillsammans med lagom mycket tur och snabba och strategiska knapptryckningar ska enligt IBM ha lett till vinsten. Förutom 1 miljondollar i prispengar och bra publicitet för IBM ser de vinsten som ett stort steg mot framtidens AI.

> *[...]* that the company has taken a big step toward a world in which intelligent machines will understand and respond to humans, and perhaps inevitably, replace some of them.
>
> *Källa: NY Times.*

Redan nu vidareutvecklas Watson's teknologi av IBM och andra företag för att eventuellt kunna ändra på hur läkare handskas med sina patienter och hur konsumenter köper produkter.

[i] DARPA betyder The Defense Advanced Research Projects Agency och är en gren i tates Department of Defense och är ansvarig för att utveckla splitter ny teknik till den amerikanska militären.

2.2 Smarta scripts



Värt att nämna är självklart ”bottar” och annan mjukvara som utvecklats av människan för att sprida spam med reklam via mail, inlägg på forum och via andra metoder. Många hemsidor gör sitt bästa för att hålla dessa datorstyrda bottar borta från sina mailservrar och hemsidor med hjälp av blockader och speciella knapptryckningar man måste göra för att bevisa att man är människa. Det kan handla om allt från att klicka in en enkel ja- eller nej-knapp som frågar ”Är du människa?” till de mer populärare metoderna med bilder med text som ligger huller om buller som en människa lätt kan skriva av. Ett dator-program kan ha väldigt svårt att läsa av dessa bokstäver och fastnar därmed på dessa säkerhetskontroller. En väldigt populär säkerhetskontroll idag som används på ett stort antal sidor kallas för ”CAPTCHA”[7] som står för ”Completely Automated Public Turing Test To Tell Computers and Humans Apart”. Vad ett Turingtest är återkommer vi till lite senare. Ett exempel på en sådan check ses i bilden här ovan. Att tolka en bild som denna kan bli väldigt svårt för en dator, men är helt klart görbart för en människa. Idag kan inte en dator läsa koden och få det korrekt nedskrivet, men uvecklingen går framåt och bot-skaparna försöker självklart få sin mjukvara att slå koden – men som CAPTCHA själva säger:

> CAPTCHA tests are based on open problems in artificial intelligence (AI): decoding images of distorted text, for instance, is well beyond the capabilities of modern computers. Therefore, CAPTCHAs also offer well-defined challenges for the AI community, and induce security researchers, as well as otherwise malicious programmers, to work on advancing the field of AI. **CAPTCHAs are thus a win-win situation: either a CAPTCHA is not broken and there is a way to differentiate humans from computers, or the CAPTCHA is broken and an AI problem is solved.**
>
> *Källa: CAPTCHAs officiella hemsida.*

2.3 Självstyrda spelare

Artificiell intelligens i dator- och konsolspel brukar man ofta kalla en teknik som går ut på att försöka imitera mänskliga spelare för att skapa en motståndare eller en lagkamrat som verkar mänsklig. Faktum är dock att termen AI oftast används för att beskriva en sammankoppling med tekniker från robotik, datorgrafik, matematik och allmän datorvetenskap.

Ett first-person shooter-spel.

AI i spel brukar vara väldigt centrerade på att ”se” bra ut. Man vill få djur och andra varelser att röra sig naturligt och få reaktionstider hos de datorstyrda spelarna att verka mänskliga. Approachen från traditionell AI är alltså lite annorlunda; man går runt problem, småfuskar och i många fall tonar ner de icke-mänskliga spelarnas förmågor för att det ska bli rättvist mot de människor som spelar. Ett exempel på detta är i ”first-person shooter”-spel där det skulle vara lätt att ge datorstyrda spelare en reaktionstid och träffsäkerhet som skulle vara totalt bortom mänsklig förmåga, men hur roligt skulle det då vara att spela?

3 Människans hjärna – en superdator?

Datorer kan göra logiska beräkningar med otroliga hastigheter och lagra minne som är felfritt och konstant utan att förändras genom tid på grund av glömska. Samtidigt är hjärnan otroligt advancerad och tillåter oss ta in mängder av information genom våra ögon och öron och tolka detta, bilda uppfattningar och ta blixtsnabba beslut. Men är datorn eller hjärnan överlägsen när man ser till processorkraft och anpassningsförmåga? Detta är en svår jämförelse att göra delvis för att hjärnan och datorn inte fungerar på samma sätt men också för att man inte vet exakt hur hjärnan fungerar.

3.1 Datorn slår hjärnan

Faktum är att hårdvaran nu börjar närma sig den mänskliga hjärnan i termen ”instruktioner per sekund” eller flops[ii]. Ray Kurzweil berättar om ”The Law of Accelerating Returns”[8] där han bygger på Moores lag[9] om datorutvecklingstrenden. Det förklaras att istället för att datorprestandad fördubblas var 18e år (Moores lag) så går det markant mer exponentiellt. Detta kan demonstreras med en tabell över utvecklingen hos superdatorer under de senaste årtiondena.

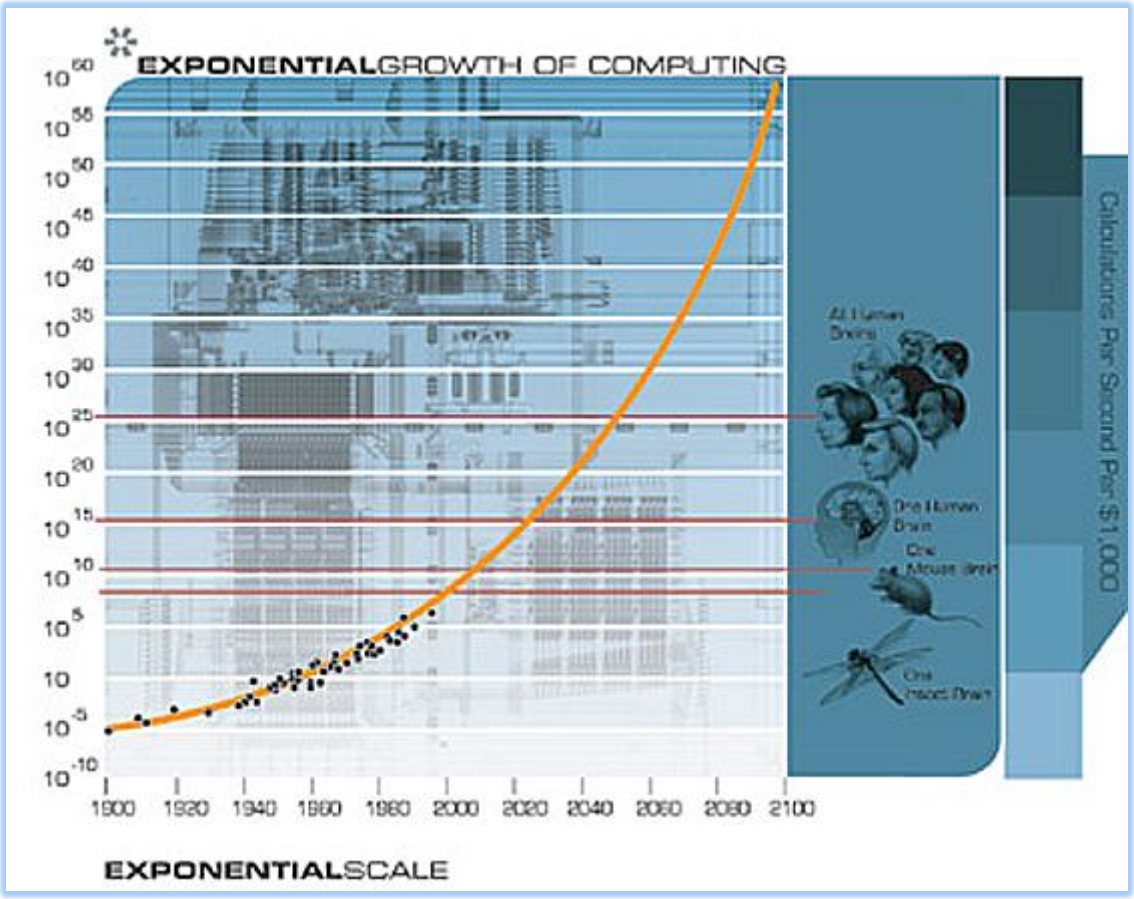
Hastighet	År	Tidsdifferans
1 Megaflops	1961	-
1 Gigaflops	1983	22 år
1 Teraflops	1997	14 år
1 Petaflops	2008	11 år

Källa: Se referens 6.

Här syns det tydligt att utvecklingen går mer exponentiellt än vad Moore förklarade. Kurzweil ger förklaringen att så fort tekniken närmar sig något sort hinder eller en barriär, så kommer ny teknik uppfinnas så att vi kan passera hindret och att detta i sin tur ger en sorts accelererande effekt. Teknik på teknik som nyttjar och är kopplade till varandra och växer likt ett oändligt stort legohus.

Kurzweil uppskattar den mänskliga hjärnan till en kapacitet på 20 petaflops $(peta - 10^{15})$[10] och än idag har ingen dator kommit upp i den hastigheten. Den idag högst uppmätta hastigheten är från en kinesisk superdator i slutet av 2010 och uppmätte 2.5 petaflops.[11] Vi ser alltså tydligt att datorn kommer gå om hjärnan om man ser till beräknings- och instruktionsbehandlingshastighet inom en snar framtid. En något äldre illustration av detta följer i bilden här under ur Kurzweils bok.

[ii] Flop betyder Floating Point Operation. Ofta används istället Flops vilket betyder Floating Operations Per Second. Flops är ett sätt att mäta en dators prestanda. http://folding.stanford.edu/English/FAQ-flops - Hämtat 29e Mars 2011.



Boken skrevs och illustrationen gjordes 2001 och visar utvecklingen fram tills dess och framtidsprognosen utifrån det. Till vänster i y-axeln har vi som bekant antal flops och längst x-axeln årtal. Till höger i figuren presenteras hjärnkapaciteten hos olika arter vilka motsvarar de röda linjerna. Från botten till toppen: En insekt, en mus, en mänsklig hjärna, alla mänskliga hjärnor kombinerat. Notera att grafen växer ganska långsamt, men ändå exponentiellt. Enligt Kurzweil har alltså skrivbordsdatorn kommit att ha samma beräknings- och processhastighet som den mänskliga hjärnan runt år 2029. Han förklarar också att runt år 2045 så kommer den artificiella intelligensen att nå en punkt där den kan förbättra *sig själv* så pass så att den frångår allting den mänskliga hjärnan någonsin skulle kunna skapa. Det som tidigare förklarades som ett oändligt legohus skulle växa så fort att vi människor hamnar långt på efterkälke. Detta är ett scenario som science fiction författaren Vernor Vinge kallar för ”Teknologisk singularitet” eller engelskans ”singularity”.

DEL II

4 Intelligenstester

4.1 Turingtest

Turingtestet introducerades år 1950 av Alan M. Turing som ett ”imitationsspel”. Syftet med testet är att avgöra om ett datorprogram är intelligent. En förklaring av originalspelet eller leken kan ges genom att citera Turing:

> The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B.
>
> *Källa: http://www.loebner.net/Prizef/TuringArticle.html*

När vi idag talar om Turingtest så är det tänkt att vår *man* och *kvinna* istället är en *människa* och en *dator*. Dessa är ansluta till någon sorts chat-program tillsammans med frågeställaren C. C’s uppdrag är att komma på vem av de två som är datorn. Datorns uppdrag är att lyckas ”lura” frågeställaren. Lyckas den så anses den vara intelligent.[12]

Testet har blivit kritiserat i en mängd olika diskussioner relaterade till AI, filosofi och kognitionsvetenskap[iii] under de senaste 60 åren. Kritiken har dessutom trappats upp ytterligare allt eftersom AIn har gjort stora framsteg inom robotiken. Kritiken har till stor det handlat om att testet inte behandlar inlärning och att det således inte behandlar intelligens. Testet behandlar konsten att imitera en människa i samtal, med allt som det kan innebära – bland annat ironi och lögn för egen vinning. Många menar på att det bara är en bråkdel av vad det innebär att vara intelligent.[13]

Vad kritiken än gäller så har Turingtestet varit ett ovärdeligt verktyg för forskningen om artificiell intelligens fram tills idag. I dagens internetsamhälle så har det bland annat utvecklats verktyg baserat på Turingtestet för att filtrera bort spam-bottar och liknande. Detta nämndes tidigare i detta dokument – bland annat när vi pratade om CAPTCHA på sida 6. CAPTCHA som står för *Completely Automated Public Turing Test To Tell Computers and Humans Apart.* Vilket är precis det testet är ämnat att göra.

[iii] Kognitionsvetenskap är ett forskningsområde när man studerar det mänskliga tänkandets natur.

4.2 Det kinesiska rummet

Det kinesiska rummet är ett tankeexpriment som John Searle introducerade år 1980 i sin artikel ”Minds, Brains and Programs”. [14] Searle utmanar Turingtestet och vill demonstrera att bara för att en maskin kan simulera en intelligent konversation betyder inte det att den nödvändigtvis förstår konversationen. Experimentet är ett av de största och mest använda motargument till att datorer en dag kommer kunna tänka.

Experimentet går ut på att en person, A, ska simulera en infödd kinesisk talare. A blir instängd i ett rum med en tjock instruktionsbok skriven på engelska. I rummet finns dessutom en stor låda fylld med kinesiska tecken. Genom en lucka i väggen skjuter människor in frågor på kinesiska och väntar sig ett svar från A. Den tjocka instruktionsboken talar om för A, utifrån den inskickade lappen, hur man kombinerar kinesiska tecken till ett svar. I boken står ingenting om tecknens betydelse. A har därför ingen aning vad varken frågorna eller svaren betyder. Människorna utanför rummet kan omöjligt veta att A inte talar kinesiska. De skickar in frågor på kinesiska och får ut vettiga svar på kinesiska. Av det drar de slutsatsen att A förstår kinesiska.

Searle menar med detta tankeexperiment att på samma sätt som A inte förstår ett ord kinesiska förstår datorn absolut ingenting men svarar ändå på användarens input. Härifrån argumenterar han vidare varför en dator inte kan vara intelligent.[15] Vi återkommer till detta i 5.2 när vi lägger fram argument mot stark AI.

5 Existensen av artificiell intelligens

5.1 Argument *för* att en dator kan vara intelligent

Det är huvudsakligen en vetenskaplig fråga att förstå människans hjärna. En stor del av denna förståelse kommer från neurovetenskap som omfattar praktiskt taget alla aspekter av hjärnan och nervsystemet, såsom perception, inlärning och minne. Stora framsteg har gjorts de senaste åren, genom positronemissionstomografi och magnetisk resonanstomografi (MRT), för att att upptäcka hur hjärnan arbetar med exempelvis språk och problemlösning.[16] Avancerade datorsystem möjliggör för neuroforskare att utforma modeller av nervceller och deras kopplingar i hjärnan. Detta arbete kan enligt Society for Neuroscience leda till datorprogram som kan förstå tal och svara på talade frågor.

Ray Kurzweil skriver i sin artikel ”Long Live AI” att det är rimligt att dra slutsatsen att vi efter två decennier kommer ha effektiva modeller för den större delen av hjärnan. För att förstå principerna för mänsklig intelligens måste vi, enligt Kurzweil, bakåtkompilera den mänskliga hjärnan. Han skriver att det har gjorts större framsteg än de flesta inser inom det området. Den spatiala och temporala resolutionen av hjärnskanning går framåt i exponentiell takt och har i stort sett fördubblats varje år. Skanningsverktyg kan nu se enskilda nervceller signalera till varandra i realtid. Vi har redan matematiska modeller och simuleringar av några dussin regioner i hjärnan, inklusive lillhjärnan, som omfattar mer än hälften av nervcellerna.[1]

Vid Institute of Electronics, Microelectronics and Nanotechnology i Frankrike har en forskargrupp ledd av Dominique Vuillaume tagit fram en ny transistor som kan komma att möjliggöra det för datorer att tänka på samma sätt som människor. Transistorn kan nämligen replikera det komplicerade sätt som nervceller, eller neuroner, i hjärnan kommunicerar med varandra via synapser. Dessa transistorer anses kunna leda till skapandet av datorer som i viss mån kan härma vårt sätt att tänka.[17]

Det har dessutom gjorts stora framsteg i kognitionsvetenskap den senaste tiden där man bygger beräkningsmodeller av mänskliga aktiviteter. Genom att konstruera modeller som kan köras på en dator kan man pröva olika typer av teorier och idéer och lära sig mer om hur människan fungerar och hur våra erfarenheter påverkar oss.[18] Om vi kan lära oss mer om den mänskliga inlärningen är det ett steg närmre att bygga en maskin med samma inlärningsmekanismer. En sådan maskin skulle förvärva kunskap och erfarenhet mycket på samma sätt som ett spädbarn.

Kognitionsvetenskapsmodellerna tillsammans med neurovetenskap kan hjälpa oss relatera det mänskliga beteendet tillbaka till våra erfarenheter och lämpliga kretsar i våra hjärnor. När vi bättre förstår våra hjärnors natur bör det inte vara omöjligt att skapa artificiella hjärnor baserade på vår förståelse.

På 1960-talet kunde den mest komplexa dator ockupera ett stort rum. Med tiden har storleken på datorer har minskat rejält och antalet kopplingselement, transistorer, har ökat. Enligt nationalencyklopedin[19] kan idag en enda integrerad krets, med stor minneskapacitet, innehålla hundramiljontals transistorer. Detta kan jämföras med hjärnans ”kopplingselement”, så kallade neuroner. Enligt Illustrerad Vetenskap finns det närmare 100 miljarder neuroner i en människas hjärna.[20]

Den mänskliga hjärnan är mindre och snabbare än dagens datorer, om man nu kan göra en sådan jämförelse. Frågan är bara hur länge detta kommer vara en sanning. Tekniken har gjort enorma framsteg och kommer med all sannolikhet fortsätta att göra det. Enligt Illustrerad Vetenskap tror många forskare att datorerna kommer gå om hjärnan redan om några decennier.

Det verkar rimligt att tro att vi genom vetenskapliga framsteg kommer göra det möjligt att förstå sinnet och genom tekniska framsteg en dag kunna bygga ett. Det finns redan idag ett stort projekt för att försöka simulera en mänsklig hjärna, som kallas “The Blue Brain Project”.[21] Projektet utgör ett viktigt första steg mot att uppnå en fullständig virtuell mänsklig hjärna och har redan demonstrerat en modell av en råttas *cortical column*[iv] som består av ungefär 10 000 nervceller. De har arbetat mycket med att förklara nervcellernas beteenden och deras sätt att ansluta till varandra och forma kretsar. Dessa iaktagelser har översatts till matematiska algoritmer för att representera nervcellernas beteende på ett realistiskt sätt.

Enligt Church Turing-tesen kan varje tänkbar beräkningsprocess utföras av en Turingmaskin. Detta argument kan även appliceras för människor och datorer. Givet ett problem som kan lösas av en människa, kan denna lösning ses som en algoritm som kan köras på en vanlig digital dator. Det är möjligt att det kommer ta mycket längre tid för datorn att lösa problemet, men om datorn har tillgång till all data som människan hade bör det vara teoretiskt möjligt.

AI-forskningsområdet grundades vid en konferens på campus Dartmouth College under sommaren 1956. Det är nu 55 år sedan. Idag kan datorer visa förståelse för mänskligt språk, lösa problem och till och med lära sig nya saker. Det kanske tar ytterligare 55 år innan vi har implementerat en intelligent maskin, det kanske tar 100 år, eller 1000 år. Poängen är att vi har all tid i världen och utveckligen går ständigt framåt.

[iv] Det finns ingen bra översättning, men en förklaring följer. Cortical Column är den del av Hjärnbarken som tros vara den del av hjärnan som är ansvarig för de högre funktionerna så som medvetande och medvetet tänkande. http://en.wikipedia.org/wiki/Cortical_column - Hämtat 12e April 2011.

5.2 Argument *mot* att en dator kan vara intelligent

Som vi beskrev i inledningen av uppsatsen är den första frågan man bör ställa sig när intelligens uppstår. Vi vill påstå att en dator, för att den ska kunna klassas som intelligent, måste vara medveten om sin egen existens. Om datorn inte vet varför den utför uppgifter, oavsett hur ”smarta” lösningar den kan hitta, är den inget annat än en beräknande maskin. ”Jag tänker, alltså är jag”, om datorn kan inse så mycket har man inte bara skapat intelligens, utan möjligtvis även liv. Nästa fråga man bör ställa sig är om detta är rimligt att åstadkomma.

Matematikern och fysikern Roger Penrose anser att hjärnan måste ha en biologisk struktur för att kunna stödja ett medvetande. Enligt hans teorier uppstår medvetandet från våglika kvantummekaniska effekter som involverar proteintrådar i nervceller. I sin bok ”Shadows of the Mind” (1997) argumenterar Penrose att den mänskliga tankeförmågan är för avancerad för att vara algoritmisk och att medvetandet måste bygga på kvantfysikaliska fenomen. Hans argument är till stor del grundade på Gödels ofullständighetsteorem.

I boken ”Att resonera logiskt” beskriver Hans Rosing Gödels ofullständighetsteorem på följande sätt:[22]

> *[...]* Det säger att i varje formellt system, som är tillräckligt komplicerat för att innehålla de naturliga talen, kan man skriva en sann formel som inte kan bevisas inom systemet.
>
> *Källa: Se referens 19.*

Penrose använder detta till att formulera ett bevis för att människans hjärna inte är en dator, genom att visa att hjärnan kan veta saker som inte ens den mest kapabla algoritm kan räkna ut. Beviset, även känt som ”Goodelian Proof”, har dock kritiserats av många matematiker, dataloger och filosofer som påpekat brister som gör att beviset inte håller.[23]

Men även om Penrose inte har lyckats med sitt bevis är det inte heller någon som lyckats bevisa motsatsen, dvs. att en hjärnan är en dator. John Searle, professor i filosofi vid University of California, hävdar också att datorer aldrig kommer bli intelligenta, även om hans åsikter skiljer sig en aning från Penroses. Searle accepterar att beteendet associerat med ett medvetande kan imiteras, men hävdar att ett sådant beteende inte innebär ett faktiskt medvetande. Penrose hävdar att vi inte ens kan få rätt beteende från en dator eftersom dess beroende av beräkning kommer att ge det begränsningar som inte delas av våra hjärnor.

Genom sitt tankeexperiment, det kinesiska rummet, som beskrevs tidigare i uppsatsen vill Searl visa att datorer inte kan vara intelligenta. Han menar på att personen i experimentet förstår Kinesiska på samma sätt en dator kan förstå språk, eller överhuvudtaget någonting den gör, med andra ord inte alls. Hans poäng är att datorer är beräknande maskiner som säkert kan efterlikna intelligens men aldrig förstå varför de gör vad de gör.[24]

Funktionalister ställer sig kritiska till Searls tankeexperiment och menar på att ett samvete kan definieras som en mängd processer inuti hjärnan. Utifrån det dras slutsatsen att allt som utför samma informationsprocesser som en människa också har ett medvetande. Med andra ord, om

vi programmerar en dator med dessa processer skulle den vara lika medvetet som en människa när den utför diverse uppgifter.

Searlie argumenterar dock, liksom Penrose, att detta är en omöjlighet eftersom medvetandet är en fysisk egenskap, såsom matsmältning eller eld. Oavsett hur bra du simulerar matsmältning så kommer ingen mat smältas och oavsett hur bra du simulerar eld kommer ingenting brännas. Observatörer plockar ut vissa mönster i världen och ser dem som informationsprocesser, men processerna i sig är inte existerande saker i världen. Eftersom de inte existerar på en fysisk nivå, hävdar Searle, kan de inte ha någon kausal effekt och därmed inte skapa ett medvetande.[25]

DEL III

6 Analys och Diskussion

6.1 Dagsläget

AI-forskare har lyckats skapa datorer som kan utföra uppgifter som är otroligt komplicerade för människor, samtidigt har de kämpat med att klara uppgifter som för en människa är helt triviala. Med dagens datorer, även om man skulle utöka prestandan enormt, anser vi att det är omöjligt att implementera en tänkande medveten maskin oavsätt hur många algoritmer man försöker proppa in. Av det vi har läst om hjärnan, tillsammans med vårt eget kunnande om datorer som programmerare och blivande civilingenjörer, drar vi slutsatsen att datorer i dagens läge inte kan jämställas med hjärnan. Datorer är strikt bestämda i sitt beteende och kan därför inte göra fria val. Det är sant att en del datorprogram kan lära sig saker av sina misstag, men det görs genom att följa algoritmer där tidigare misstag eller händelser sparas i minnet och kan återanvändas på ett eller annat sätt. Det är sant att datorer kan vinna över schackspelare på världsnivå, men det beror på datorernas enorma beräkningskapacitet som den mänskliga hjärnan inte kan jämföras med. Det har ingenting med intelligens att göra. Lika lite som att en miniräknare är intelligent för att den kan beräkna avancerade multiplikationer på millisekunder som för en människa skulle ta mycket lång tid.

Att Turingtest fortfarande används idag kan man vara kritisk mot. Många professorer och forskare har länge varit kritiska – framför allt då för att de själva anser att en enkel konversation inte avgör om en dator är intelligent eller inte. Det testet gör är att utgå ifrån faktumet att vår intelligens är baserad på hur vi beter oss och utifrån det kunna avgöra vad som är intelligent. Som vi alla dock vet är detta inte sant – för man kan alltid säga eller göra en sak men mena en annan. Med detta inkluderas bland annat misstag, lögn och ironi.

Om man nu frågar en dator vad den *tycker* om vädret så kommer man inte få en åsikt, detta eftersom en dator inte har en åsikt. Den har svar, minne och hastighet. Artighet och ironi kan alltid programmeras in smart – men en dator kan idag inte ge *sin* ärliga åsikt för den tycker helt enkelt ingenting. Den har inget medvetande. Professorerna anser därför att Turingtestet är vänligt i ett internetsamhälle men skrapar knappt på ytan av vad intelligens egentligen är.[26]

Det kinesiska rummet är däremot en vidare mer accepterad term. Detta av anledningen att argumentet går mycket djupare ner i de bakomliggande faktorerna. Argumentet förklarar bland annat för oss att det *finns mer till medvetandet än vårt beteende*.

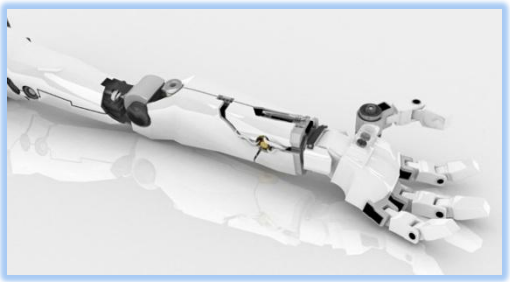
6.2 Vi ser framåt

Vem säger att inte datorerna kommer se annorlunda ut i framtiden? I argumenten för stark AI nämns försök i Frankrike att replikera sättet neuroner i hjärnan kommunicerar med varandra. Blue Brain Project forskar för att simulera en mänsklig hjärna, samtidigt finns det säkert fler forskningsprojekt inom området världen över. Kanske kommer man kunna få datorerna att mer och mer efterlikna hjärnans sätt att hantera information.

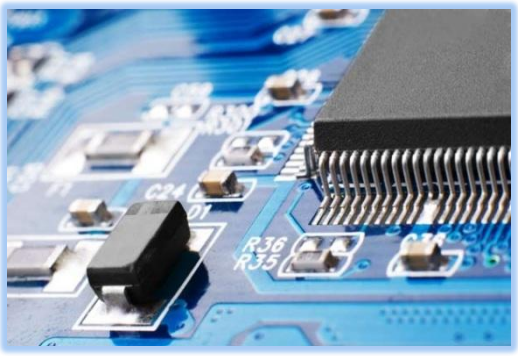
Om Roger Penrose och John Searle har rätt i sina teorier om att människans medvetande måste bygga på kvantfysikaliska fenomen är det möjligt att detta aldrig kommer gå att simulera med digitala komponenter. I framtiden kanske man kan skapa intelligens genom att använda både biologiska och digitala komponenter. Det låter i dagsläget mycket som science fiction, men människan är uppfinningsrik och tenderar att hela tiden göra det ”omöjliga” möjligt. Ett exempel på detta sätter Alan Winfield, professor på University of the West of England, som för närvarande forskar med hybridrobotar. I hans forskningsprojekt provar han att sätta hjärnceller från råttor i robotar, så att bara hjärnan på roboten är biologisk medan resten är mekaniskt. Genom att göra detta kan man jämföra de markanta skillnaderna mellan dessa hybridrobotar och robotar med ”mjukvaruhjärnor”. Winfield kommer bland annat fram till att den biologiska hjärnan ändrar sin fysiska uppbyggnad under inlärning och blir således mycket bra på det den lär sig, men det tar istället tid. [27]

Det ultimata målet med artificiell intelligensen vore att skapa ett medvetande. Men hur avgör vi om vi har lyckats? Det är en grundläggande fråga för AI som leder in på såkallad solopism: en idé inom filosofi att förekomsten av medvetande i en annan varelse inte kan bevisas. Men om medvetandet inte kan bevisas, kan ingenjörer och forskare någonsin visa att AI har skapats? Självklart underlättar det om roboten kan prata och kommunicera och därigenom skulle man kunna dra slutsatser. Det är samma problem som att avgöra om djur som katter eller apor har ett medvetande. Vet djuren om att de existerar, kan de tänka fritt? Eller går de genom livet och lever på hundra procent instinkter och inlärning?

Att ha en hjärna går hand i hand med att ha en kropp, men är det ett krav? Forskare inom robotik är högst beroende av AI-forskning eftersom de behöver AI för att få deras robotar att fungera. Men är det omvända sant även för AI-forskarna? En del tror att man inte kan befria hjärnan från kroppen, att det måste ses som en helhet med hela kroppen eftersom hjärnan inte arbetar isolerat. Hjärnan fungerar tillsammans med immunsystemet och det endokrina systemet och det autonoma nervsystemet - liksom sig självt och det centrala nervsystemet. Samtidigt är många AI-forskare inte alls intresserade av robotar eftersom de tror att de kan implementera ett intellekt direkt i en dator. [28]



Om intelligens inte behöver en kropp kanske vi i framtiden kommer ha nätverk som kan tänka, tycka och övervaka sig själva. Eller kylskåp som tänker och tycker att du äter för lite nyttig mat? För att undvika att sväva ut för mycket så sätter vi punkt för diskussionen kring intelligens och kroppar här.



Det finns många som tror att det bara är en tidsfråga tills vi får AI som kan jämföras med det mänskliga intellektet och ofta hänvisar man till Moore's lag. Men denna lag om teknikutvecklingens takt sätter inte en lag på att allt är tekniskt möjligt. Bara för att vi får datorer som teoretiskt sett är tillräckligt snabba för att simulera en hjärna betyder inte det att det faktiskt är möjligt att utföra. Till exempel kan man inte bygga en bro enbart för att man har tillräckligt stora mängder stål. Bara för att vi kan fantisera om ett resultat så betyder inte det att det någonsin kommer att vara möjligt. Fler exempel på detta skulle vara fordon som svävar och därmed trotsar gravitationen eller möjligheten att kunna uppnå total osynlighet genom perfekt omdirigering av ljusstrålar.

Det är alltid svårt att förutse framtiden. Om man kunde åka tillbaka några hundra år i tiden och ställa frågan om människan någonsin kommer kunna landa på månen skulle de flesta förmodligen svara att det är omöjligt. Kanske är det samma sak med AI. När vi försöker förutse framtiden utgår vi ofta från den teknologin vi har idag, men morgondagen kommer ständigt med nya upptäckter.

> People who are much wiser than myself have made predictions about the future that have turned out to be ridiculously false. I think it's not because they were stupid, it's because such predictions are impossible to make.
>
> *Källa: Daphne Koller, professor inom AI i Stanford University in California. [24]*

7 Slutsats

Du kan lära en dator otroligt mycket genom att stoppa in algoritm efter algoritm. Teoretiskt sett skulle en dator med oslagbar prestanda efter oändligt många algoritmer kunna allt förutom att tänka. Ett samvete är inte något som magiskt uppstår bara för att man lär en maskin tillräckligt mycket. Likadant är intelligens inte ett mått på hur många uppgifter man kan utföra. Intelligens är ens förmågan att tänka – att existera. Vi tror inte att det kan beskrivas med algoritmer. Med det säger vi inte att medvetandet är något magiskt som står utanför vetenskapens gränser. Någonting sker i hjärnan som gör oss medvetna, oavsett om det beror på kvantfysikaliska fenomen eller något annat.

Det högsta man kan hoppas åstadkomma med algoritmiska processer är att efterlikna intelligens, aldrig att skapa det. En dator kan hitta bästa vägen från norra Stockholm till södra Stockholm på kortast tid, en annan kan köra bil igenom en storstad på egen hand och en tredje kan snabbt hitta ketchupen längst in i matvaruhuset. Det är imponerande men det är inte intelligens, det behövs någonting mer.

Detta är slutsatsen vi drar, som mycket väl kan vara fel. I framtiden kanske datorer kommer se helt annorlunda ut och kunskapen om hjärnan vara häpnandsväckande mycket större. En dag kanske vi kommer förstå exakt vad samvetet är och hur det skapas, även digitalt. Men om den dagen kommer är det förmodligen långt efter vår levnadstid.

Referenslista

[1] http://www.forbes.com/home/free_forbes/2005/0815/030.html - Hämtat 14:e juni 2011. Artikel om AI av Ray Kurzweil.
[2] http://en.wikipedia.org/wiki/IBM_Deep_Blue - Hämtat 31a Mars 2011. Schacklösningssystemet.
[3] http://archive.darpa.mil/grandchallenge05/ - Hämtat 31a Mars 2011. DARPA Grand Challenge 2005.
[4] http://en.wikipedia.org/wiki/DARPA_Grand_Challenge_%282005%29 – Hämtat 31a Mars 2011. Information om tävlingen gällande resultat.
[5] http://www.silicon.com/technology/software/2010/02/08/artificial-intelligence-55-years-of-research-later-and-where-is-ai-now-39503564/4/ - Hämtat 13e April 2011. DARPA 2007, storstadstrafik.
[6] http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?_r=2 – Hämtat 31a Mars 2011. Artikel om Jeopardy-datorn ”Watson”.
[7] http://www.captcha.net/ - Hämtat 2a April 2011. CAPTCHAs officiella hemsida.
[8]http://www.kurzweilai.net/the-law-of-accelerating-returns - Hämtat 29e Mars 2011. Ray Kurzweils bok ”The Law of Accelerating Returns”. Illustrationer och information.
[9] ftp://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf - Hämtat 29e Mars 2011. Moores lag.
[10] http://illuminati.wordpress.com/2007/11/13/2015-a-machine-that-thinks-as-fast-as-the-human-brain/ - Hämtat 29e Mars 2011. Idéer, illustrationer och länkar.
[11] http://www.bbc.co.uk/news/technology-11644252 - Hämtat 29e Mars 2011. Kinesisk superdator satte nytt rekord.
[12] http://www.fil.ion.ucl.ac.uk/~asaygin/tt/ttest.html - Hämtat 12e April 2011. Kort info om Turing Test.
[13] http://www.v3.co.uk/v3-uk/opinion/1995525/is-turing-test-retire - Hämtat 12e April 2011. Argument mot Turingtest.
[14] http://www.iep.utm.edu/chineser/ - Hämtat 12e April 2011. Information om Kinesiska Rummet.
[15] http://www.iep.utm.edu/chineser/#H1 – Hämtat 12e April 2011. Kinesiska experimentet i detalj.
[16] http://www.sfn.org/index.aspx?pagename=whatisneuroscience – Hämtat 10e April 2011. Information om människans nervsystem och vetenskapen kring detta.
[17] http://mobil.idg.se/2.1085/1.302873 - Hämtat 9e April 2011. Transistorer härmar den mänskliga hjärnans neuroner.
[18] http://sv.wikipedia.org/wiki/Kognitionsvetenskap - Hämtat 10e April 2011. Kort information om Kognition.
[19] http://www.ne.se/integrerad-krets) – Hämtat 9e April 2011. Nationalencyklopedin om integrerade kretsar.
[20] http://illvet.se/fraga-oss/har-hjarnan-storre-kapacitet-an-en-superdator - Hämtat 11e April 2011. Antalet neuroner i den mänskliga hjärnan samt informationen till de nästkommande styckena.
[21] http://bluebrain.epfl.ch/ – Hämtat 10e April 2011. The Blue Brain Project’s officiella hemsida.
[22] http://web.abo.fi/fak/hf/filosofi/HRlogik/kapitel10.pdf - Hämtat 12e April 2011. Hans Rosing, Att resonera logiskt, kap 10.
[23] http://www.paul-almond.com/RefutationofPenroseGodelTuring.htm - Hämtat 12e April 2011. Penrose’s bevis.
[24] http://www.iep.utm.edu/chineser/ - Hämtat 12e April 2011. Kinesiska rummet.
[25] http://plato.stanford.edu/entries/chinese-room/ - Hämtat 12e April 2011. Kinesiska rummet.
[26] http://www.tjonard.ws/chinese.html - Hämtat 14e April 2011. Stöd till diskussion.
[27] http://www.silicon.com/technology/software/2010/02/08/artificial-intelligence-55-years-of-research-later-and-where-is-ai-now-39503564/2/ - Hämtat 13e April 2011. Stöd för diskussion.
[28] http://www.silicon.com/technology/software/2010/02/04/artificial-intelligence-can-ai-crack-the-conundrum-of-consciousness-39503567/2/ - Hämtat 13e April 2011. Stöd för diskussion.