

## Indholdsfortegnelse

Indledning.....	2
Problemformulering .....	2
Metode.....	2
Kildekritik.....	2
KI.....	3
Introduktion til KI.....	3
Det klassiske symbolparadigme.....	5
Ingeniør AI: .....	5
Psykologisk AI: .....	5
Gruppe A: .....	5
Gruppe B: .....	6
Funktionalismen .....	6
Klassisk AI i praksis.....	7
GPS.....	7
ELIZA & SHRDLU .....	7
Deep Blue.....	8
Turing-testen .....	8
Diskussion af Klassisk AI.....	9
5 trin fra nybegynder til ekspert .....	10
Hvad fortæller modellen? .....	12
Mennesket vs. Maskinen .....	12
Kroppen og Common Sense.....	13
China-Box.....	14
Opsummering .....	15
Det neurale netværks paradigme .....	15
Perspektiv hos det kunstige neurale netværk .....	17
Konklusion .....	18
Litteraturliste.....	19

## Indledning

Kunstig intelligens er både spændende, skræmmende, fascinerende og uhyggelig på samme tid, fordi den sætter menneskets unikhed i perspektiv. I science fiction film, som "AI" og "iRobot" forstilles det, at fremtidig teknologi og forskning kan give robotter samme kognitive egenskaber som mennesket. Men hvor langt er det fra virkeligheden? *"Der er altid en aktuel – og teknologisk bestemt – grænse for, hvilke kognitive præstationer som vi i dag kan "løsrive" fra mennesket og eksternalisere i ydre fysiske systemer og maskiner, men grænsen flytter sig sammen med den teknologiske udvikling."*<sup>1</sup> Problemstillingen er spændende i og med, at teknologi vinder mere og mere indpas i vores dagligdag, og ligeledes vil kunstig intelligens følge med i mere eller mindre avanceret form. Tanken om at kunne stille mere præcise diagnoser, sparre tonsvis af resurser og i det hele taget forbedre verden på mange områder, har ført og vil fortsat føre forskningen frem i søgelyset.

## Problemformulering

*Der ønskes en redegørelse af kunstig intelligens inden for det klassiske symbolparadigme. Efterfulgt ønskes en diskussion om, hvorvidt klassisk AI holdbarhed i forhold til at opnå kunstigt intelligens på niveau med menneskelige eksperter er mulig? Til sidst ønskes der en perspektivering til et moderne paradigme.*

## Metode

Opgaven vil være inddelt i tre overordnede afsnit. Første del, vil være en præsentation og redegørelse af klassisk AI ved indbringelse af forskellige teorier og grene. Anden del vil tage form som en diskussion, bestående af en kritik af den klassiske AI, hvor Hubert Dreyfus i form af hans 5-trins læringsmodel og hans tillægelse af kroppens betydning vil blive inddraget sammen med John Searles tankeeksperiment.. Tredje og sidste del af opgaven vil være en dybdegående perspektivering, som vil have fokus på det mere moderne AI paradigme, nemlig konnektionisme eller de såkaldte kunstige neurale netværk. Opgaven vil blive rundet af med en konklusion, som vil belyse problemformuleringen.

## Kildekritik

Den udvalgte litteratur til at beskrive klassisk AI og tilhørende kritik, synes fyldestgørende, da det i mange sammenhænge stadig fungerer som primærlitteratur. Dog er en del af værkerne passeret de 25 år, og eventuelle sammenligninger med dengang nutidens computere er måske ikke længere troværdige.

---

<sup>1</sup> Wackerhausen (1989): s.121

Mht. intelligensbegrebets definition, vil jeg støtte mig op ad Steen Wackerhausens tilgang, hvor det ikke kun er et fænomen, som kan tilgås via levende væsner, men derimod som en intellektuel præstation, som maskiner har samme vilkår for at opnå<sup>2</sup>. Jeg er opmærksom på at forskellige forfattere kan have forskellige forståelser, og dette vil blive belyst, hvis det skulle forekomme.

Litteraturen vedrørende de kunstige neurale netværk har været svært at få i hænde, men jeg anser det anvendte materiale, som fuldt ud anvendelig til en perspektivering.

## KI

Indenfor kunstig intelligens (fremover KI), kan der i grove træk overordnet tales om tre paradigmer, der hver især har deres syn på, hvad tænkning er og hvordan intelligens kan skabes kunstigt. Først er det *klassiske symbolparadigme*, som er dominerende fra 1950 og frem. Dette bliver efterfulgt af *konnektionismen* eller de såkaldte *neurale netværk*. Tredje paradigme er *embodied KI*, og faktisk er der i dag tale om et helt fjerde paradigme, som fokuserer på selvorganisering. Denne opgave vil dog kun tage udgangspunkt i de to førstnævnte.

Før en gennemgang af det klassiske symbolparadigme, vil jeg dog først undersøge baggrunden og opståen af KI.

## Introduktion til KI

Begrebet kunstig intelligens<sup>3</sup> (KI) dækker over maskiner, der i stand til at overveje, lære og tage beslutninger på samme niveau som mennesket. Udtrykket blev første gang anvendt af John McCarthy i en ansøgning om midler til en konference i Darmouth<sup>4</sup> i 1955, som var forbundet med computerens fremtræden i 1940'erne og 50'erne. Deltagende i denne messe, var nogle af de førende forskere indenfor området bl.a. Allen Newell og Herbert Simon<sup>5</sup>, som grundlagde et af de mest indflydelsesrige laboratorier indenfor KI.

Drømmen om at mekanisere den menneskelige tænkning, stammer faktisk helt tilbage fra den græske filosof Aristoteles, der grundlagde den formelle logik. Det var her ikke begrebernes betydning, som var i fokus, men derimod den overordnede logiske sammenhæng, der dannes af ord som fx "og",

---

<sup>2</sup> Wackerhausen (1989): s. 115

<sup>3</sup> Kunstig intelligens er en direkte oversættelse af det engelske *Artificial Intelligence* (AI). Jeg bruger både forkortelser KI og AI, og de skal fortolkes ens!

<sup>4</sup> Darmouth konferencen anses som den første af sin slags og bliver betragtet som dén historiske begivenhed, der gjorde KI til et selvstændigt forskningsområde.

<sup>5</sup> Disse vil der blive vendt tilbage til senere i afsnittet om klassisk AI i praksis.

"eller" etc. Hans tilgang var dog forholdsvis begrænset, men i 1800- og 1900-tallet blev den formelle logik udviklet væsentlig af bl.a. George Boole<sup>6</sup>. Han var grundlægger af den såkaldte binære logik, som i dag stadig udgør hjertet i alle computere – bl.a. ved brug af de såkaldte boolske operatorer, såsom "and" og "or". Inden for den filosofiske tradition begynder ligeledes en særlig tankegang, hvor det forestilles, at alle tanker kan udtrykkes som symboler, og tænkning foregår som en ren mekanisk procedure, hvor symbolerne håndteres efter bestemte formelle regler. Samtidig med Boole, opstår behovet for mekaniske regnemaskiner. Pascal havde dog allerede i 1645 lavet en *pascalina*<sup>7</sup>. Brugen var dog begrænset, da den ikke var programmerbar. Dette problem bliver dog belyst af Charles Babbage<sup>8</sup> i 1837, hvor han præsenterer ideen om en programmerbar maskine, der ved brug af samme fysiske hardware, kunne løse forskellige problemer afhængig af, hvilket program, der var aktivt. Dette var eksempelvis de populære hulkortmaskiner, som bl.a. blev anvendt indenfor vævningsindustrien. Hans såkaldte *Analytical Engine*, blev dog aldrig fuldført, men blev inspiration til de første elektroniske computere, der i 1940'erne blev de første programmerbare beregningsmaskiner<sup>9</sup>.

Kombinationen af den programmerbare maskine og den rationelle opfattelse af, at tænkning består af en mekanisk beregning, resulterede i de første forsøg på at skabe tænkende maskiner. Grundlaget for KI var lagt, og det i næstkommende afsnit vil det klassiske symbolparadigme blive gennemgået, som netop tager udgangspunkt i denne kombination.

Førend der kan fortsættes til de egentlige paradigmer, skal der først afklares et par ting. KI inddeles nemlig i tre overordnede grundantagelser, uafhængigt af hvilket paradigme man opererer indenfor. De tre grundsætninger er som følgende:

1. Grundantagelse: *"Menneskets kognitive færdigheder og præstationer kan "læsres" fra mennesket og eftergøres computationelt."*<sup>10</sup>
2. Grundantagelse: *"Menneskets indre kognitive processer er af en computationel karakter"*<sup>11</sup>
3. Grundantagelse: *"De emotionelle, bevidsthedsmæssige fænomener er også computationelle funktionelle fænomener."*<sup>12</sup>

---

<sup>6</sup> George Boole (1815-1864): Britisk matematiker og filosof. Hans udarbejdelse af den boolske algebra, blev og er fundamentet i alle moderne computere, og som har givet ham titlen som værende fader til *Computer Science* (da: datalogi), trods computeren ikke eksisterede i hans levetid.

<sup>7</sup> Pascalina er en maskine der kan håndtere de fire regnearter plus, minus, gange og divider.

<sup>8</sup> Charles Babbage (1771-1871): Engelsk filosof, matematiker og ingeniør.

<sup>9</sup> Homoartefakt.dk artikel B

<sup>10</sup> Wackerhausen (1989): s.112

<sup>11</sup> Wackerhausen (1989): s.112

<sup>12</sup> Wackerhausen (1989): s.112

Vi vil i det næste afsnit se hvordan forskellige grene indenfor klassisk AI, arbejder ud fra de tre ovenforstående grundantagelser.

## Det klassiske symbolparadigme

Det klassiske symbolparadigme opstod som tidligere nævnt på baggrund af kombinationen af ny teknologi og en ny rationel tankegang i forbindelse med menneskets kognitive processer. Det gælder, at al tænkning er mekaniske operationer i et formelt system. Indenfor det klassiske symbolparadigme, findes dog overordnet nogle forskellige grene, som hver især tager udgangspunkt i hver deres grundantagelse.

### Ingeniør AI:

Med udgangspunkt i første grundantagelse, fokuserer ingeniør AI primært på at skabe computermaskiner, der kan udføre "intelligente operationer" for os. Med det menes, at målet er en forbedring af menneskets kognitive og intellektuelle færdigheder såsom regnekraft, problemløsning, ekspertviden etc. Den er resultatorienteret og fokuserer på input-output. Der er ikke nær så meget fokus på, hvordan processen egentlig fungerer i menneskets hjerne, og det er heller ikke ønsket at genskabe en nøjagtig kopi af processen, men derimod kun af resultatet. Ingeniør AI er altså resultatorienteret og har kun i sinde at opfylde første grundantagelse. Her kommer psykologisk AI ind i billedet<sup>13</sup>.

### Psykologisk AI:

Denne gren har modsat ingeniør AI fokus på beskrivelse og forståelse af de indre mentale processer og fænomener, der ligger til grund for menneskes kognitive færdigheder. Dvs. de mellemliggende processer fx mellem input-output ækvivalensen. Her er dog uenighed blandt psykologisk AI, hvilket derfor opdeles i to grupper A og B, som hver især tager udgangspunkt i henholdsvis 2.- og 3.grundantagelse<sup>14</sup>.

### Gruppe A:

Gruppe A bygger videre på 1.grundantagelse og mener, som allerede nævnt, at de indre kognitive processer er mulige at genskabe som KI, men derimod ikke de emotionelle og bevidsthedsmæssige. De indre kognitive processer er altså af en computational karakter og dermed fuldt ud beskrivbare. Der er altså her tale om, at gruppe A sammenligner menneskets mentale "inderside" med en maskine, hvad enten det

---

<sup>13</sup> Wackerhausen (1989): s. 112

<sup>14</sup> Wackerhausen (1989): s. 112

er den serielle von Neumann type<sup>15</sup>, eller den af neural-netværks typen. "Kort sagt, så er budskabet hos Psykologisk AI", at vi kan komme til en korrekt forståelse af menneskets psykiske "inderside" ved at se mennesket i computerens billede!"<sup>16</sup>

### Gruppe B:

Gruppe B bygger her videre på de to ovenstående antagelser, men mener altså også, at alt i den menneskelige hjerne (herunder de emotionelle og bevidsthedsmæssige processer) er eksternaliserbare til computere med den rette software.

Nøglebegrebet indenfor klassisk AI er computation<sup>17</sup>. De kognitive processer, som konstituerer menneskets intelligens og som muliggør menneskets intellektuelle præstationer, blev og bliver stadig opfattet som værende af en computationel størrelse. Begrebet computation er defineret som formelle manipulationer af abstrakte symboler på grundlag af eksplicite formelle regler. Klassisk AI ser kognition som regelstyret symbolmanipulation. Ifølge Newell & Simon så er intelligens, "*The Ability to store and manipulate symbols*", og de kognitive processer er syntaktisk producerede frembringelser af sekvenser af symbolske udtryk. For at opsummere, så er de nødvendige og tilstrækkelige betingelser for, at et fysisk system kan udvise en generel intelligens, at det er et fysisk symbolmanipulerende system, hvilket både repræsenteres af hjernen såvel som computeren. Dette betyder eksplicit at hvis menneskets kognition ifølge klassisk AI består i regelstyret symbolmanipulation, så er menneskets kognition fuldt ud beskrivbar, og heraf følger også programmerbar i forbindelse med computerbaseret KI.<sup>18</sup>

### Funktionalismen<sup>19</sup>

Funktionalismen opfylder den 3. Grundsætning. Denne opfattelse er på mange punkter klassisk AIs officielle bevidsthedsfilosofi. Funktionalismen mener med udgangspunkt i Klassisk AI og datalogi (computer science), at tilmed de emotionelle fænomener inklusiv de bevidsthedsmæssige og følelsesmæssige er at forstå som relationelle funktionelle størrelse. "At være i en bestemt tilstand (som er realiseret i noget "hardware" af en eller anden slags, fx en hjerne, men potentielt også en computer)". At være i bestemt emotionel tilstand, er altså det samme som at være i en funktionel tilstand. Ligesom at det samme computerprogram kan køre forskellige hardware, så er realiseringen af en mental tilstand ikke bundet til én type "hardware". Det

<sup>15</sup> Forstås som en almindelig computer med én processor. Brunak & Lauptrup (1988): s.73-76

<sup>16</sup> Wackerhausen (1989): s. 118

<sup>17</sup> Wackerhausen (1989): s. 121

<sup>18</sup> Wackerhausen (1989): s. 122

<sup>19</sup> Wackerhausen (1989): s. 122

væsentlige er dog om en given "hardware" i virkeligheden kan realisere de mentale tilstande, hvilket gruppe B mener, at moderne computere med den korrekte software kan. *"If functionalism is correct, it follows not only that scientific theories of mentality can be simulated by computer programs, but also that in principle mentality can be embodied within an "appropriately programmed computer"*".<sup>20</sup> Man kan altså sige, hvad bevidstheden er for hjernen, er programmet for computeren.

## Klassisk AI i praksis

I dette afsnit vil der blive præsenteret nogle af de maskiner og systemer, som blev produceret/programmeret under det klassiske symbolparadigme.

### GPS

A. Newell og H. Simon spillede, som tidligere nævnt, en stor rolle inden for KI. De var ansat i RAND Corporation og var sammen med Cliff Shaw blandt de første forskere, der forsøgte at forstå virkelighedens genstande som symboler, eksempelvis ord eller skakbrættet. I et af deres største projekter, ville de undersøge hvilken metode studerende anvendte under problemløsning, for derigennem at implementere det i et system. De opfandt termen "heuristisk program" for ikke at forveksle det med de såkaldte algoritmiske programmer, som systematisk fulgte en metode. De fandt frem til en række generelle metoder til problemløsning og fik dem i 1957 implementeret i et system – *General Problem Solver* (heraf GPS), som viste sig at kunne løse mange typer af problemer. Teamet siger: *"Intuition, indsigt og stor viden er ikke længere noget, som kun menneskelige væsener har; enhver stor og hurtig computer kan programmeres til også at besidde disse egenskaber."*<sup>21</sup> Dog kom H. Dreyfus hurtigt med kritik i en efterfølgende rapport, hvor han belyste, at trods de enkelte succesoplevelser man havde haft, ikke skulle sætte for høje forhåbninger.

### ELIZA & SHRDLU

Op gennem 1960'erne og 1970'erne blomstrer mange projekter frem. ELIZA, som med moderne et udtryk er en chatrobot, blev skrevet af datalogiprofessoren Joseph Weizenbaum. Han fik udstyret programmet med nogle semantiske evner, og som ved hjælp af nogle simple tricks kunne udvise en intelligent adfærd. Fx kunne man sige "Jeg føler mig trist tilpas", hvorefter den ville svare "hvorfor føler De Dem skidt tilpas?".

---

<sup>20</sup> Wackerhausen (1989): s. 123

<sup>21</sup> Dreyfus (1986): s. 24

Hvis ikke den havde nogen typisk reaktion kunne den sige: "Fortæl mig om deres far." Folk lod sig narre af denne adfærd og mange brugere blotlagde deres dybeste følelser over for en maskine<sup>22</sup>.

Det vidste sig dog imidlertid, at programmet havde mange fejl og indgik fx "nej" i inputtet, så svarede den "lad være med at være så negativ", da den var programmeret til det, hvis "nej" indgik. Terry Winograd kom dog et skridt nærmere en generel KI, med sin SHRDLU. SHRDLU simulerer på en tv-skærm med en robotarm, der kan flytte nogle klodser af forskellig form, men gør det samtidigt også muligt for brugeren at indgå i en dialog med computeren, ved at stille spørgsmål og give kommandoer inden for den begrænsede verden med klodser. Den indeholdt ligeledes grammatiske regler, semantik, men også fakta om klodser. Man kunne fx spørge "Tag en stor, rød klods op", hvor efterfølgende den sagde ok, og simulerede det på skærmen. Effekten var imponerende og programmet udviste, hvad virkede som en generel intelligens indenfor dens mikroverden<sup>23</sup>.

### Deep Blue

Et mere moderne eksempel er IBMs skakcomputer Deep Blue. En specielfremstillet supercomputer med fokus på at slå verdens bedste skakspiller Garry Kasparov. Første duel var i 1989, hvor Kasparov uden problemer vandt en to sæts kamp. Andet mødet var i 1996 med en forbedret udgave, hvor Garry tabte første kamp. Dette fik ham til at tage maskinen seriøst, og efter at have analyseret spillet vandt han overlegen de næste tre kampe. Deres tredje og sidste duel var 3.maj 1997, hvor den forbedrede Deep Blue, tabte første match, men endte med at vinde med 3½ point mod Kasparovs 2½. Efterfølgende beskyldte Garry IBM for at have brugt ufine metoder, men resultatet talte for sig selv, nemlig at maskinen var mennesket overlegen i en intelligent disciplin.<sup>24</sup>

### Turing-testen

Turing-testen er et redskab til at afgøre, hvorvidt en maskine udviser intelligens på samme niveau som et menneske, uden at være nødsaget til at definere præcis, hvad menneskelig intelligens egentlig er for en størrelse. Testen blev foreslået af den engelske matematiker Alan M. Turing i artiklen *Computing Machinery and Intelligence* fra 1950. Turing-testen tager udgangspunkt i spillet *The Imitation Game*, som går ud på om en person ud fra to personer kan gætte, hvem der henholdsvis er kvinden eller manden. Dette opnås ved at stille spørgsmål via papirform, hvorefter manden og kvinden prøver at overbevise testpersonerne, at de er

---

<sup>22</sup> Dreyfus (1986): s. 114

<sup>23</sup> Dreyfus (1986): s. 115-118

<sup>24</sup> Homoartefakt.dk artikel A

af det modsatte køn. I Turing-testen stiller Turing spørgsmålet "*What will happen when a machine takes the part of A in the game?*"<sup>25</sup> Vil testpersonerne være i stand til at svare rigtigt lige så ofte som havde det været en mand og en kvinde, hvilket vil føre til hans første spørgsmål "*can machines think?*"<sup>26</sup> I forsøget på at finde svaret fører det ham til ni indvendinger mod den kritik, som spørgsmålet indtil 1950 havde stødt på. De vil ikke blive gennemgået i denne opgave, men eksempelvis kan næves kritikken fra Professor Geoffrey Jefferson, om at maskiner ikke kan tænke, da de ikke har en bevidsthed. Jefferson mener, at maskiner først kan sammenlignes med menneskets hjerne, når de kan skrive en sonet eller en koncert grundet deres tanker og følelser. Turing svarer hertil, at man ikke kan vide om andre end en selv rent faktisk oplever følelser, og derfor kan turing-testen ikke afvises. Han fortsætter, "*I do not wish to give the impression that I think there is no mystery about consciousness (...) but I do not think these mysteries necessarily need to be solved before we can answer the question..*"<sup>27</sup> Denne problemstilling vil blive omtalt i diskussionen med inddragelse af John Searles *Chinese Room*.

## Diskussion af Klassisk AI

Dette afsnit har til opgave at stille klassisk AI i et mere kritisk søgelys, og undersøge, hvor holdbar klassisk Als tilgang til at skabe kunstig intelligente er. Dette gøres ved inddragelse af anerkendte og indflydelsesrige kritikere, som mener at hjernen og hele mennesket er mere komplekst ind blot et computabelt stykke software, som klassisk AI foreskriver. Hovedvægten i denne diskussion vil ligge på Hubert L. Dreyfus og hans bror Stuarts værk, *Intuitiv ekspertise. Den bristede drøm om tænkende maskiner*<sup>28</sup>, hvor bl.a. deres 5.trins ekspertlæringsmodel vil være i fokus. Ligeledes vil jeg bruge et afsnit fra Hubert Dreyfus' værk *What Computers can't do. The limits of Artificial Intelligence*. Her vil jeg inddrage afsnittet om kroppens rolle i forbindelse med intelligens. Derudover vil jeg også inddrage filosofen John Searle og hans china-box eksperiment. Først vil jeg dog først redegøre for Dreyfus' argument for at kunstige systemer aldrig vil opnå ekspertviden på niveau med mennesket, nemlig deres 5 trins læringsmodel.

---

<sup>25</sup> Turing (1950): s.434

<sup>26</sup> Turing (1950): s.433

<sup>27</sup> Turing (1950): s.447

<sup>28</sup> Bogens danske titel. Original titel: *Mind over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*.

## 5 trin fra nybegynder til ekspert

Dreyfus argumenterer for, at man ikke kan afvise klassisk AI uden at vide, hvordan mennesket egentlig selv opnår en ekspertpræstation. Deres metode har været at undersøge en række forskellige discipliner, bl.a. skak, sygeplejen, køreskoler etc., for derigennem at finde fælles tendenser, som kunne resultere i en generel model. Jeg anser det ligeledes nødvendigt at inddrage denne model, dog vil jeg gøre det i en stiliseret udgave, hvor jeg kun vil bruge eksempler fra deres skak-undersøgelse. Desuden mener Dreyfus ikke, at menneskets kognition skal forstås som regelstyrede symbolarbejdningsprocesser, dvs ikke-computationelle, og slet ikke når der er tale om ekspert-præstationer<sup>29</sup>.

At handle intuitivt er knowhow<sup>30</sup>. At handle regelbaseret er knowthat. Det er to vigtige begreber, som Dreyfus bruger gennem i sin 5 trinsmodel. At blive ekspert er en overgang fra at handle regelbaseret til at være intuitiv. Denne overgang vil nu blive gennemgået.

1. **Nybegynder**<sup>31</sup>: Som nybegynder lærer man gennem undervisning at genkende forskellige objektive kendsgerninger og træk, der er relevante, for derigennem at lære de tilstedeværende regler. Elementerne i situationen bliver så klart defineret, at de tydeligt forstås uden for konteksten, og disse kaldes kontekstfrie. Ligeledes er reglerne også kontekstfrie. En nybegynder i skak, lærer således at spille skak efter en formel. Man lærer at hver brik har én værdi, og at man minimum udbytter brikker med samme værdi hos modstanderen.
2. **Avanceret nybegynder**<sup>32</sup>: På dette stadie har man fået betydeligt mere erfaring med virkelige situationer. Gennem praktisk erfaring i konkrete situationer, begynder den avancerede nybegynder at genkende elementer i situationen. Disse elementer kaldes situationsbestemte, for ikke at forveksle dem med de kontekstfrie elementer. Reglerne er altså nu baseret på både de situationsbestemte og de kontekstfrie komponenter. Som skakspiller lærer man på dette stadie at genkende og undgå for spredte stillinger. Samtidig kan man genkende mange situationsbestemte aspekter ved en stilling, fx kan man se, at ens åbning har medført en svag kongeside. Dog kan man stadig ikke definere universalt gældende regler.

---

<sup>29</sup> Wackerhausen (1989): s.125

<sup>30</sup> Brødrene Dreyfus skriver om intuition og knowhow: *"Intuition eller knowhow, som vi forstår det, er hverken vildt gætterier eller overnaturlig inspiration, men en evne, som vi alle bruger hele tiden, når vi passer vore daglige gøremål.* Dreyfus (1986): s. 55

<sup>31</sup> Dreyfus (1986): s. 44-45

<sup>32</sup> Dreyfus (1986): s.46-47

3. **Kompetent**<sup>33</sup>: Som kompetent udøver bliver antallet af genkendelige kontekstfrie og situationsbetingede elementer fra den virkelige verden voldsomt forøget. Der benyttes en hierarkisk procedure, når der skal træffes beslutninger. Først vælges en plan, hvorefter faktorer undersøges for derigennem at forenkle og forbedre sin præstation. Den kompetente er mål-fokuseret og det ses også, at udøveren føler sig involveret i at nå målet. I skak kan en spiller efter at studeret en angrebsplan nøje vælge at angribe modstanderen. Han ser bort fra eventuelle svagheder ved den nye position og angrebet bliver det altoverskyggende mål.
4. **Kyndige**<sup>34</sup>: Den kyndige udøver er dybt involveret i sin opgave. Der sker ikke noget objektivt valg eller nogen objektiv overvejelse – det sker bare. Den kyndige har nemlig erfaringer med lignende situationer fra tidligere lejligheder, og derfor associerer de nuværende situationer med planer, der tidligere har virket, og forudser resultater, der tidligere har vist sig. Evnen til intuitivt at reagere på mønstre uden at opløse dem i deres bestanddele kaldes "holistisk skelnen og association"<sup>35</sup>. Samtidigt med man handler intuitivt, så tænkes der alligevel analytisk på, hvad der skal gøres. En kyndig skakspiller kan genkende et meget stort antal af stillinger. Han ved, uden at skulle anstrenge sig, hurtigt, hvad en stilling indebærer og kan således ved beregning derfra fortsætte sin intuitive plan. Fx ved han, at der skal angribes, men må overveje og analyserer frem, hvordan det gøres bedst.
5. **Ekspertise**<sup>36</sup>: En ekspert *ved*, hvad han skal gøre på basis af indsigt og erfaring. Dybt engageret i sin opgave tænker man ikke længere over ens handlinger. Ens færdighed er blevet en del af en selv. En stormester i skak ser ikke længere spillet som 32 brikker på et bræt, men derimod som en krig, hvor man selv er en involveret aktør i en verden med trusler, styrker, svagheder og muligheder. En ekspert skal dog ikke ses som det idealiserede billede, at eksperter aldrig tænker og altid har ret. En ekspert overvejer også en situation, hvis tiden tillader det. Det drejer sig dog ikke om kalkulatorisk problemløsning, men snarere kritisk refleksion over intuitionen. Evnen til at skelne mellem et stort antal situationer kommer med erfaringen, og man gemmer rent faktisk flere situationer i hjernen end man gemmer ord. Stormesteren kan desuden straks ud fra en bestemt stilling finde det næste træk, og faktisk kan en ekspert spille med en hastighed på 5-10 sek. per træk uden at spille væsentligt dårligere. Dermed er det også sagt, at præstationen udelukkende er afhængig af intuition og mindre analyse.

---

<sup>33</sup> Dreyfus (1986): s. 47-50

<sup>34</sup> Dreyfus (1986): s. 53-56

<sup>35</sup> Dreyfus (1986): s. 54

<sup>36</sup> Dreyfus (1986): s. 57-62

## Hvad fortæller modellen?

Det burde fremgå klart at progressionen fra den analytiske adfærd, som uengagerede personer udviser, samt opløse sine omgivelser i genkendelige elementer og følge abstrakte regler, til den involverede adfærd, som er baseret på en holistisk gruppering af nye situationer over for de dermed forbudne reaktioner, hvilket skyldes heldige erfaringer med lignede situationer. Dreyfus konkluderer: *"Intelligens er mere end kalkulatorisk fornuftstænkning. Skønt irrationel adfærd – dvs. adfærd, der strider mod logikken eller fornuften – normalt bør undgås, følger det ikke deraf, at det at optræde rationelt bør betragtes som det endelige mål*<sup>37</sup>. Med andre ord, så findes der altså et område mellem irrationelle og det rationelle som kan kaldes *arationel*. *"Kompetent adfærd er rationel; kyndig adfærd er et mellemstadium; eksperter handler arationelt"*.<sup>38</sup>

## Mennesket vs. Maskinen

Der er lavet forsøg med spillet tetris hvor man fx kan forestille sig, hvordan man drejer en puslespilsbrik for at se om den passer, frem for at lukke øjnene og beregne. Med andre ord løser mennesket beregningsopgaven ved at handle og sanse, hvilket computeren ikke har mulighed for.

Når vi kigger på computerens præcision, pålidelighed og utrættelighed, når de bruges som logiske maskiner<sup>39</sup>, så er de på mange områder mennesket overlegen. Uanset hvor disciplinær man gør mennesker, så vil de aldrig opnå samme præcision som de regelbundne maskiner.<sup>40</sup> Til gengæld udviser mennesker en fleksibilitet, en dømmekraft og en intuition, som ikke lader sig opløse i specifikationer og logiske følgeslutninger jævnfør de fem trin. Dette har vidst sig at være vanskeligt at påføre de logiske maskiner, selv med den bedste programmering, så spørgsmålet er hvilket niveau kan maskinen opnå?

Computere som logiske maskiner er ideelle begyndere. De bliver en slags ekspert indenfor begyndere, eftersom deres viden er fuldkommen, når de har fået kodet regler og hundredvis af aspekter ind. De behøver altså ikke engang at øve sig. Når programmøren forsøger at få computeren videre til næste skridt, nemlig den avancerede begynder, opstår der problemer, da det her kræver, at man kan genkende tilbagevendende betydningsfulde elementer. Nyere forskning viser at sammenligningen mellem den nye og tidligere erfaret situation foregår som en holistisk billedmekanisme, og vil derfor ikke være muligt at genskabe i en logisk maskine. De kan dog programmeres til at organisere deres kontekstfrie kendsgerninger under hensyntagen til mål som et kompetent menneske. Computeren kan nemlig få en regel, som fortæller

---

<sup>37</sup> Dreyfus (1986): s. 65

<sup>38</sup> Dreyfus (1986): s. 65

<sup>39</sup> Dreyfus (1986): s. 87-94

<sup>40</sup> Dreyfus (1986): s. 103

den, at hvis visse kendsgerninger er til stede, bør situationen organiseres under hensyntagen til bestemte mål.<sup>41</sup> Den vil dog ifølge Dreyfus aldrig nå op på de sidste to færdighedsniveauer, da det her mennesket handler ubevidst intentionelt og intuitivt.

### Kroppen og Common Sense

Som vi har set, så viser færdighedsmodellen, at menneskets kognitive aktiviteter og evner ikke er regelstyrede processer, i hvert fald kun en lille del deraf. Selv indenfor skak, hvor man ellers kunne formode det, da der er klare spilleregler i et velstruktureret universers. Her handler eksperten efter intuitionen, modsat amatøreren, der eksplicit tænker logisk og rationelt. Selv med indbringelse af computeren i skak, kan den ikke præ-udregne alle spillemuligheder eller brikpositioner. Der opstår en kombinatorisk eksplosion som selv verdens hurtigste computer ikke kan vil kunne håndtere. Men ved at udstyre computeren med samme "heuristiske" regler, som en ekspert anvender, så kan computere måske få succes. Dog har de maskiner, såsom Deep Blue, som har haft størst succes, været baseret på "rå" regnekraft, hvilket helt klart ikke er menneskets metode.

Det er ingen tvivl om, at Deep Blue var succesfuld og om ikke bedre, så jævnbyrdig med den bedste intuitive ekspert. Men dette er ikke nødvendigvis ensbetydende med, at det vil sprede sig til alle ekspertområder. Det gælder nemlig for skak, at det er et utroligt velstrukturerede universers med klare præcise regler, og al nødvendig viden om spillet er eksplicit i form af, at brikkerne er synlige. Andre ekspertområder såsom medicin og geologi, da klare ekspertsystemer sig langt fra så godt.<sup>42</sup>

Vi så i projektet SHRDLU, at det rent faktisk var muligt at kommunikere med en maskine i et mikro-univers. Men hvordan vil klassisk AI klare sig, når domænet ikke længere er afgrænset til et specifikt område? Computeren bliver her nødt til at få regler for hvilke facts, der relevante i en given situation, for hvis den skal gennemgå alle facts, vil den brænde sammen eller være for langsom. Dette er ikke problem for mennesket, hvor begrebet *common sense* eller sund fornuft tager sig af det. Dreyfus beskriver i *What computers can't do*, at mange af menneskets kognitive færdigheder og *common sense* er knyttet til det forhold, at mennesker er engagerede og intentionelle. Mennesket har nemlig en biologisk krop og derigennem mål, drifter og behov og "*an involved, situated, material body*".<sup>43</sup> Dette aspekt har computeren ikke og siden den ikke befinder sig i en situation, må den behandle alle facts, som altid værende mulige<sup>44</sup>. Derimod skriver Dreyfus om mennesket at dets verden er: "*prestructured in terms of human purposes and*

---

<sup>41</sup> Dreyfus (1986): s. 105

<sup>42</sup> Dreyfus (1986): s. 125

<sup>43</sup> Dreyfus (1979): s. 230

<sup>44</sup> Dreyfus (1979): s. 258

*concerns in such a way that what counts as an object or is significant about an object already is a function of, or embodies that concern (...). This cannot be matched by a computer, which can deal only with universally definable, i.e. context-free, objects".*<sup>45</sup> Dreyfus afviser også den 3. grundantagelse, for selvom man kan simulere menneskelige behov på en computer, så vil det ikke løse problemet, da denne tilgang blot vil gøre emotionerne og deres effekter til "objekter". Dreyfus mener netop, det er fordi vores emotioner ikke er *"objects in our experience that they can play this fundamental role of organizing our experience into meaningful patterns or regions."*<sup>46</sup>

### China-Box

Den amerikanske filosof og AI kritiker John Searle, har en lidt anden tilgang til klassisk AI, idet han opdeler den i *svag* og *stærk* KI. Svag AI er systemer, der blot er i stand til at efterligne visse træk ved menneskelig tænkning, mens stærk KI er systemer, der virkelig har intelligens af samme karakter og dybde som mennesker<sup>47</sup>. Hans kritik forholder sig til KI af den stærke og art, men også den tidligere omtalte Turing-test. Searle præsenterede i 1980 et tankeeksperiment, som kan beskrives som en stor kasse, med et hul til input og et til output, hvor der henholdsvis kan komme kinesiske skrifttegn ind og ud. Inde i kassen er en række kinesiske skrifttegn i en kurv, samt en stor bog, hvor der står hvilke skrifttegn, der skal bruges som output til hvert input. Bogen står på engelsk og kan sammenlignes med algoritme/computerprogram. Der placeres nu en englænder i kassen, som hverken kan skrive eller tale kinesisk. Systemet startes ved at give et kinesisk tegn og ved at bruge den engelske manual finder han det rigtige kinesiske svar og sender det ud som output.<sup>48</sup>

Systemet virker intelligent udefra, da det kan svare på de spørgsmål, der bliver stillet, men lukkes englænderen ud af kassen ses det, at han stadigvæk ikke forstår kinesisk, og at han heller ikke har forstået noget, imens han var inde i kassen. En ydre input-output ækvivalens med mennesket er altså ingen garanti for en indre ækvivalens. Ifølge Searle, så skal der en biologisk hjerne til for, at intentionalitet og mentale tilstande kan opstå.<sup>49</sup> Steen Wackerhausen konkluderer på *china-boxen* at, *"i min opfattelse lykkes Searles argument dog kun på eet punkt, nemlig som et bevis for at input-output ækvivalens ikke er en garanti for tilstedeværelsen af den indre ækvivalens, og dermed som et bevis for Turing-testens utroværdighed.*

---

<sup>45</sup> Dreyfus (1979): s. 261

<sup>46</sup> Dreyfus (1979): s. 274

<sup>47</sup> Searle (1981): s. 282-283

<sup>48</sup> Searle (1981): s. 284-286

<sup>49</sup> Wackerhausen (1989): s. 129

## Opsummering

Da dette er en diskussion er det ikke nødvendigt at finde et konkret svar på, hvor vidt om den klassiske tilgang er holdbar eller ej. Ligesom med mange andre videnskabelige områder, så er det også naturligt at stille spørgsmålstejn og derigennem opnå mere præcise definitioner. Jeg har valgt nogle anerkendte kritikere frem, som begge er enige om, at et computerprogram ikke vil kunne opnå en intelligens og bevidsthed svarende til menneskets. De fleste tilgange til at skabe KI har været gennem problemstillinger, men hertil siger Dreyfus: *"Vi er enige i, at problemløsning er "tilstrækkelig" til at frembringe visse former for intelligent adfærd; det er veldokumenteret. Men der er ikke gnist af bevis for, at det er "nødvendigt", at vi ikke kan være intelligente uden at løse problemer."* Vejen synes altså endnu lang for at skabe en generel KI.

## Det neurale netværks paradigme

Indtil nu er klassisk AI og en tilhørende diskussion blevet gennemgået. Dette afsnit har ikke til opgave at svare på alle de kritiske spørgsmål, der blev omtalt i diskussionen, men derimod at se KI fra en anden vinkel. Der er altså tale om en perspektivering, som ikke vil blive udsat for en diskussion, men blot åbne nye døre til forståelsen og mulighederne af KI.

Et af alternativerne til den klassiske AI findes hos Moderne AI. Tilgangen til Moderne AI foregår gennem det neurale netværks-paradigme. Med det menes, at man prøver at lave en præcis simulering af hjernen gennem et kunstigt neuralt netværk, bestående af kunstige neuroner og tilhørende synapser. Først der kan fortsættes vil der kort blive redegjort for menneskets hjerne.

Menneskets hjerne fylder cirka halvanden liter, hvor størstedelen er bestående af vand. Hjernen består af en hvid masse, der er omgivet af hjernebarken, som er ansvarlig for hjernens fænomenele informationsbearbejdende egenskaber.<sup>50</sup> Hjernebarken er cirka 3 millimeter tyk og bestående af 100 milliarder neuroner, som alle er forbundet med hinanden i et kompliceret ledningsnet. Hver neuron har sin egen identitet, men mødes med andre neuroner i bestemte kontaktpunkter, nemlig synapser. Herigennem kan de overføre signal til hinanden, og de er typisk forbundet med mellem 1000-10.000 andre neuroner. Regnekapaciteten er enorm og den samlede ækvivalente regnekapacitet er omkring  $10^{18}$  operationer per sekund. *"Dette er mere end 100 millioner gange hurtigere, end nogen computer kan gøre det i dag".*<sup>51</sup>

I et kunstigt neuralt netværk (KNN) ønsker man at skabe samme opbygning som menneskets hjerne. For at beskrive neuronernes aktivitetsniveau bruges en simplificeret tilgang, hvor en neuron enten er aktiv eller inaktiv. På den måde kan man aflæse hele netværkets aktivitetstilstand på én gang. Ved at

<sup>50</sup> Brunak & Lauptrup (1988): s. 29

<sup>51</sup> I dag betyder her 1988, da dette er bogens udgivelsesdato, og sammenligningen bør evt. revurderes. Brunak & Lauptrup (1988): s. 33

tildele den aktive tilstand værdien 1 og den inaktive værdien 0, kan man i fx det mest simple neurale netværk bestående af fem neuroner, aflæse en værdi på 01101 eller 11001. Her er der altså ingen forskel på neurale netværks aktivitetsmønstre og den logiske tilstand af en digital computer. Forskellen ligger i den indre dynamik, som vi vil se nærmere på nu.

Kommunikationen foregår gennem synapserne og neuronernes aktivitetsniveau bestemmer om indflydelsen er høj eller lav på andre neuroner. Det er de kunstige synaptiske styrker, som tilsammen bestemmer netværkets indre dynamik. Neuronerne afgør løbende om de skal være aktive eller inaktive på baggrund af den samlede indflydelse – *det kollegiale pres*.<sup>52</sup> Selvom neuronerne er selvstændige kan de blive nødt til at bukke under for det denne indflydelse. Presset vil hele tiden sørge for at netværket stiler mod en stabil tilstand. Modsat en computer, som behandler alting i en synkron rækkefølge, så fungerer det neurale netværk som en asynkron computer, hvor det ikke spiller nogen afgørende rolle, om de enkelte neuroner er ude af takt. Netværkets dynamiske opførsel er robust over for taktløshed<sup>53</sup>.

Et kunstigt neuralt netværk bliver først interessant som computer, når det bliver i stand til at løse veldefinerede beregningsopgaver. Den grundlæggende forskel er, at systemet ikke skal programmeres som et stykke software fra klassisk AI, men derimod *oplæres* og lære alting fra bunden (lige som menneskets hjerne). Gennem et *ydre pres* kan man få netværket til at skifte aktivitetsniveau, for til sidst at ramme det rigtige aktivitetsniveau. Som vi tidligere har hørt, så er menneskets hukommelse baseret på associationer til de erindringer den indeholder, og det er på den måde, vi fx kan genkende en ellers glemt vits ved blot at høre starten af den. Dette, som ellers er en flaskehals for den klassiske tilgang, kan optrænes i det neurale netværk. Netværket gives et input i form af et aktivitetsmønster, der repræsenterer en klump information, som fx kunne være stumperne af et indlært bogstav. Neuronerne udkæmper, parallelt og kollektivt, en kamp, der til sidst vil hensætte det i en stabil tilstand med minimal frustration, nemlig den fuldstændige erindring. De kunstige synapsers styrker afgør, hvilke aktivitetsmønstre, som er stabile, og bestemmer altså, hvad modellen er i stand til at huske<sup>54</sup>. Der foregår altså to forskellige frustrationsformindskende processer. Den som nævnt ovenfor ved association, og den anden ved indlæringsproces aktivitetsmønstret justeres indtil det er korrekt.

Det vil være for omfattende i denne opgave at fremvise de beregninger som bl.a. bruges til oplæring, men fx kan der nævnes en anerkendt indlæringsregel *Delta Rule*, som ved hjælp af komplekse

---

<sup>52</sup> Brunak & Lauptrup (1988): s. 91

<sup>53</sup> Brunak & Lauptrup (1988): s. 94

<sup>54</sup> Brunak & Lauptrup (1988): s. 96-97

regnestykker hurtigt kan stimulere og stabilisere det samlede aktivitetsmønster i kunstigt neuralt netværk<sup>55</sup>.

## Perspektiv hos det kunstige neurale netværk

Livet som vi kender det er blevet til gennem én ud af mange mulige evolutionslinjer. Med det menes at kombinationen af mange forskellige fænomener har gjort at verden ser ud som den gør i dag. Var Jorden fx ikke blevet ramt af en meteor for 65 millioner år siden, var evolutionslinjen anderledes. For biologer er dette et interessant område at forske i, og KNN kan her være behjælpelig, da de ligesom levende systemer kan tilegne sig flere forskellige opgaver i flere forskellige miljøer. Man kan altså simulere naturlige organismer. En organisms nervesystem kan således repræsenteres ved et KNN på en computer. Der benyttes såkaldte *genetiske algoritmer*<sup>56</sup>, som man kan give forskellige overlevelseskriterier. KNN kan, som tidligere nævnt, repræsenteres i form af en bit-streng i en computer, og denne udgør det KNNs genotype. Den genetiske algoritme udvælger et antal genotype, som skal bruges til reproduktion ifølge det opstillede overlevelseskriterium. Gennem reproduktion konstrueres generation på generation, og derved har man skabt en hel evolutionsproces. Gennem denne tilgang, også kaldet *Artificial Life*, kan man studere naturen som den kunne være opstået. Det er dog stadig usikkert om man vil opnå systemer, som vil opleve følelser som os, og have en sjæl og bevidsthed, men gennem *Artificial Life* kan disse spørgsmål testes og verificeres<sup>57</sup>.

---

<sup>55</sup> Berkeley (1997)

<sup>56</sup> Genetisk algoritme simulerer på computeren naturlige processer så som reproduktion, crossover og mutation af gener ud fra et darwinistisk synspunkt med survival of the fittest.

<sup>57</sup> Hauptop (1995)

## Konklusion

Tilgangen til at skabe KI sker i dag gennem flere paradigmer, men først og måske vigtigst er det klassiske symbolparadigme. Det er logisk og naturligt at tænke i den serielle sekventielle computer, da det i mange henseende også er så mennesket fungerer. For os er det umuligt at læse 7 bøger på én gang og ordene følger efter hinanden, ét efter ét. Resultaterne ser i dag ikke ud af meget, men datiden taget i betragtning, da var de epokegørende, dog så meget, at de gik fra *"store forhåbninger til barsk virkelighed"*.

Det er klart at sådanne nye tiltag vil blive massiv kritiseret. Sådan vil det altid være indenfor videnskaben. Især, når man stiller menneskets unikhed i perspektiv, vil der opstå en masse kritiske spørgsmål, da mennesket nærmest føler sig truet som første led i fødekæden. Spørgsmålet om en fuld eksternalisering af menneskets kognitive færdigheder er mulig, eller endda en forbedring af menneskes intelligens, kan ikke besvares. Rent empirisk vides det endnu ikke, om det er principperne, eller blot den nuværende teknologi, som bremser udviklingen. Det kan derfor hverken bekræftes eller afkræftes, om denne klassiske tilgang er forældet.

Det ubesvarede spørgsmål har dog ikke bremset udviklingen, og tilgangen via de moderne paradigmer har vidst sig at være utroligt interessant og er begyndt at få megen opmærksomhed. KNN kan med en meget præcis og korrekt simulation af hjernen, nemlig tilpasse sig skiftende miljøer, da den er oplært og ikke regelbunden, og den holistiske og associative skelnen som mennesker bruger til at genkende elementer, har gode vilkår i sådant et system. Hvorvidt denne tilgang med tiden vil opfylde 3. grundantagelse vil jeg lade stå ubesvaret hen og blot lade empirien tale for selv.

## Litteraturliste

### Primær litteratur

- Brunak, Søren & Lautrup, Benny (1988): *Neurale Netværk. Computere med intuition*. Munksgaard.
- Drefys, Hubert & Dreyfus, Stuart (1986): *Intuiv Ekspertise. Den bristede drøm om tænkende maskiner*. Munksgaard.
- Dreyfus, Hubert (1979): *What Computers can't do. The limits of artificial intelligence*. Harber & Row. New York
- Searle, John (1981): *Minds, Brains, and Programs*. Redigeret af John Haugeland 1981. MIT Press
- Turing, Alan (1950): *Computing, Machinery and Intelligence*. Mind, årg. 59, Nr. 236
- Wackerhausen, Steen (1989): *Mennesket i computerens billede*. Philosophia, årg. 18.

### Artikler

- Hauptop, Henrik Lund (1995): *Livet som det kunne være*. Weekendavisen, 24. Feb. – 2.marts.
- Berkeley, Istvan S. N. (1997): *What is Connectionism*.  
<http://www.ucslouisiana.edu/~isb9112/dept/phil341/wisconn.html>

### Websider

- [www.Homoartefakt.dk](http://www.Homoartefakt.dk)
  - Atikler:
    - A: *Teknologien i menneskets udvikling*.
    - B: *Kunstig intelligens*