

The nature of intelligence: can man re-create himself?

Een essay over de vraag of intelligentie na te bouwen is

Inleiding

In dit essay wordt er ingegaan op de vraag of het, in de toekomst, mogelijk is om systemen te ontwikkelen welke dezelfde (of zelfs meer) intelligentie bezitten dan mensen.

De afgelopen jaren is er in de films “I, Robot” en “AI” een beeld neergezet van onze toekomstige wereld waarin robots zich op een gelijk of zelfs hoger intelligentie niveau bevinden dan de mensheid. In deze films zijn de robots daardoor superieur aan de mensheid en ontstaat zodoende het beeld alsof robots een volgende stap in de evolutie zouden zijn. Deze films roepen zodoende onder andere de volgende emoties c.q. vragen op: “is deze situatie in de toekomst echt mogelijk, of blijft het bij fictie? En willen we deze situatie wel?”

Onze samenleving ontwikkelt zich, in de laatste decennia, steeds stormachtiger. Zo moeten organisaties steeds sneller betere producten en/of diensten leveren aan hun afzetmarkt en willen consumenten meer hulpmiddelen in hun dagelijkse leven om het leven eenvoudiger c.q. gemakkelijker te maken.

Aangezien ICT ons kan helpen om deze ontwikkelingen bij te houden, wordt er al jaren intensief onderzoek gedaan op het gebied van *artificial intelligence*, het terrein waarin de praktijk uit de genoemde films gerealiseerd tracht te worden. Dit heeft tot doel om ICT te ontwikkelen welke hetzelfde intelligentie niveau kan bereiken als mensen, waardoor deze ICT ons kan helpen op het gebied van veranderingen, efficiëntere dienstverlening e.d. ICT is echter een zeer breed begrip. Zoals te zien is in de documentaire van Discovery Channel “Robosapiens” [Discovery] kost het al zeer veel moeite om een robot te laten lopen. Naar aanleiding van de definitie van *intelligence* zoals deze in dit essay wordt gehanteerd, is het niet nodig om zulke aspecten (het kunnen lopen) aan bod te laten komen. In dit essay richt ik mij daarom op intelligente softwaresystemen, en niet op specifiek op de hardware. Zo gaat het er niet om of het mogelijk is om robots te laten voortbewegen, want is dit wel intelligentie?

De (ethische) vraag die in dit essay beantwoordt zal worden is dan ook de volgende: “Is het nu, of in de toekomst, mogelijk om software te ontwikkelen welke net zo intelligent is als de mens? Zo ja, wat voor ethische en maatschappelijke veranderingen kunnen er dan verwacht worden?”

Op het gebied van de kunstmatige intelligentie zijn er veel verschillende stromingen, definities e.d. in omloop. Om het bereik van deze vraag enigszins in te perken zal er in de komende sectie een positionering worden gegeven van de belangrijkste begrippen. In het vervolg van dit essay zal er een korte samenvatting gegeven worden van de ontwikkelingen op dit gebied tot op dit moment en de verwachte toekomstige ontwikkelingen. Daarna kan er antwoord gegeven worden op de zojuist gestelde vraag en de ethische aspecten hiervan.

Positionering van begrippen

Zoals, al gezegd zijn er zeer veel begrippen en definities van deze begrippen in omloop. In deze sectie zullen de belangrijkste begrippen worden gedefinieerd.

Intelligentie

Zoals gezegd, is er niet één geaccepteerde definitie van intelligentie. Om te bepalen of het mogelijk is om intelligente software te ontwikkelen moet er toch een definitie gehanteerd worden.

Henk Tuten [Tuten] definieert intelligentie als volgt: “*Intelligence is random associating of units of knowledge using some kind of grammar called language*”. FreeDictionary [FD]: “*The capacity to acquire and apply knowledge, the faculty of thought and reason and superior powers of mind*”. Syracuse University [Syracuse]: “*Ability to deal with abstract concepts and form complex pictures of the outside world. Creativity, ability with spoken and written language etc.*” Wikipedia [Wikipedia]: “*Intelligence is a general mental capability that involves the ability to reason, plan, solve problems, think abstractly, comprehend ideas and language, and learn*”.

De definitie van Wikipedia past het best bij de gestelde vraag. Immers, wil software voldoen aan onze intelligentie standaard, moet het deze eigenschappen bezitten. De eerste drie definities schieten zodoende te kort en zijn niet specifiek genoeg.

Hierbij moet echter aangetekend worden dat ook het begrip *consciousness* een belangrijke rol speelt bij de intelligentie. Veel onderzoekers geven aan 'iets' alleen intelligent kan zijn als het ook bewust is van zichzelf en de omgeving. In de rest van dit essay zal daarom de definitie van Wikipedia met de toevoeging *consciousness* gehanteerd worden.

Kunstmatige intelligentie

Nu intelligentie gedefinieerd is, is het mogelijk om kunstmatige intelligentie te beschrijven. John McCarthy [McCarthy], grondlegger van dit vakgebied: "*It is the science and engineering of making intelligent machines, especially intelligent computer programs*". AIII [AIII]: "*the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines*".

Deze definities zijn uitbreidingen op de definitie welke gegeven is door Luca Consoli [Consoli]: "*the scientific effort to build and design intelligent artifacts*".

Over dit begrip heerst minder verwarring. Het gaat er weldegelijk om om objecten te creëren welke intelligent zijn.

Weak AI vs. Strong AI

Er zijn een hoop verschillende definities van het begrip intelligentie. Dit is mede te wijten aan het feit dat er eigenlijk op twee terreinen onderzoek gedaan wordt naar intelligentie, namelijk de *weak AI* en *strong AI*.

Weak AI is gebaseerd op computers welke problemen kunnen oplossen in een gelimiteerd kennisdomein, maar geen echte intelligentie tentoon kunnen spreiden [Wikipedia]. Met behulp van de Turing test [Syracuse,AIII,Copeland] is het mogelijk om computer te testen op deze criteria. Op dit terrein is er al veel vooruitgang geboekt en worden ook als succesvol beschouwd. Echter, wij mensen kunnen meer! John Searle heeft deze, hogere, vorm van kunstmatige intelligentie *strong AI* genoemd [Searle]: "*according to strong 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. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations.*"

Strong AI is niets meer en niets minder dan het streven naar het creëren van een systeem welke precies gelijk is aan het menselijke brein, zodat ware kunstmatige intelligentie behaald kan worden. Een dergelijk systeem is dan ook bewust van zijn aanwezigheid. Jape Copeland [Copeland] heeft beschreven dat computers / software intelligent zijn als ze kunnen leren, kunnen denken, problemen kunnen oplossen, kunnen waarnemen en taal kunnen herkennen.

Artificial Intelligent Agents & viable systems

We willen dus toe naar software / hardware / robots welke zelfstandig moet kunnen leren, denken, etc. De *agents* technologie neemt in dit striven een prominente rol in. Agents kunnen als volgt gedefinieerd worden: "*a hardware or (more usually) software-based computer system that enjoys the properties of autonomy, social ability, reactivity, pro-activeness*" [Wooldridge, Jennings].

Een agent heeft dus de capaciteit om zelfstandig te opereren binnen een omgeving, te communiceren met andere agents, de omgeving zelfstandig waar te nemen en op eigen initiatief kunnen handelen (proactief reguleren). Een self-adaptive agent kan dan ook gedefinieerd worden als een agent met autonome, adaptieve, samenwerkende, flexibele en lerende eigenschappen.

Willen we in de toekomst toe naar intelligente machines, is de agenttechnologie een *must*, tevens omdat met agents *viable systems* gecreëerd kunnen worden [Buitenhuis]. Stafford Beer [Beer] heeft op onderzoek gedaan naar de benodigde systemen waardoor mensen en andere organismen levensvatbaar zijn. Beer beschrijft de levensvatbaarheid van systemen als volgt: "*viability is maintained by engaging in different activities, keeping them from interfering with each other, managing them together, focusing on the future and doing so in the context of an identity within which the interests of the whole over time could be considered*".

In deze definitie zien we grote raakvlakken met de *AI* definities; met andere woorden: intelligente systemen zijn levensvatbare systemen.

Neurale netwerken

Nu gedefinieerd is dat agents belangrijk zijn in de zoektocht naar kunstmatige intelligentie, is het zinvol om de techniek te beschrijven waarmee dit alles gerealiseerd moet worden. Na jarenlang onderzoek is gebleken dat de menselijke intelligentie zich hoofdzakelijk bevindt in ons brein [Syracuse]. Men is daarom gaan zoeken naar een oplossing welke het brein kan nabootsen, zodat intelligente systemen ontwikkeld kunnen worden. *Neurale netwerken* bezitten deze capaciteit.

Neurale netwerken zijn ontworpen naar analogie van het biologische brein, m.a.w. een zeer uitgebreid en onderling verbonden netwerk van neuronen. Kunstmatige neurale netwerken zijn toe te passen in computersystemen. Een kunstmatig neuraal netwerk bestaat uit verscheidene (eventueel door software gesimuleerde) meestal zeer eenvoudige processoren (neuronen) met een hoge mate van onderlinge connectie waarover simpele scalaire berichten verzonden worden. De interactie tussen de diverse onderling verbonden processoren, waaruit het netwerk bestaat is bovendien adaptief, zodat verbindingen tussen andere processoren in het neurale netwerk kunnen ontstaan, kunnen worden versterkt, verzwakt of weer verbroken kunnen worden. Dit betekent dat een neuraal netwerk te 'trainen' is.[Wikipedia]

Probleem op dit moment is echter dat deze techniek in de praktijk nog niet uitontwikkeld is, op dit moment hebben kunstmatige neurale netwerken de intelligentie van een slak [Vuik] en is er een langdurig, *supervised learning process* nodig om tot acceptabele resultaten te komen [Syracuse].

Stergiou en Siganos [Imperial College London] hebben een paper gepubliceerd waarin wordt geconcludeerd dat: *“The most exciting aspect of neural networks is the possibility that some day 'conscious' networks might be produced. There is a number of scientists arguing that consciousness is a 'mechanical' property and that 'conscious' neural networks are a realistic possibility. ... even though neural networks have a huge potential we will only get the best of them when they are intergrated with computing, AI, fuzzy logic and related subjects.”*

Feit is echter wel de fundamenteën van deze techniek uitstekend zijn om *strong AI* mee te kunnen ontwikkelen. Immers, deze techniek biedt de mogelijkheid om “te leren door te leren” en heeft geen *rule-based* principe [Syracuse]. Hierdoor kan een *agent* met een *neuraal netwerk*, in principe, meer leren dan de ontwikkelaar heeft geprogrammeerd. Tot op heden is dit de meest veelbelovende techniek welke beschikbaar is.

Ontwikkelingen tot nu

Om antwoord te kunnen geven op de gestelde vraag is het noodzakelijk om te weten wat er tot op heden bereikt is op het gebied van de kunstmatige intelligentie. Tevens biedt deze sectie de mogelijkheid om in het kort weer te geven hoe de stromingen binnen de *AI* wereld gevormd zijn.

Voor 1950

Het is een misverstand als gedacht wordt dat het onderzoek naar kunstmatige intelligentie vanaf de jaren 50 van de vorige eeuw begonnen is. Hephaestus en Pygmalion zijn in de oudheid al bezig geweest met intelligente robots, heeft onder andere Hero bijgedragen aan het construeren van mechanische robots en heeft Aristoteles het eerste deductie systeem ontwikkeld [Jonathan Cohen].

In de 15^e tot de 19^e eeuw is er hoofdzakelijk onderzoek gedaan naar het framework van het denken, de eerste cybernetische theorieën waarin mensen als machines werden beschreven en, uiteraard, het vervaardigen van mechanische rekenmachines [Buchanan].

In het begin van de 20^{ste} eeuw zijn nieuwe stappen gezet op het gebied van de formele logica [Russel & Whitehead], de cybernetica [Ashby], zijn de eerste testen ontwikkeld om kunstmatige intelligentie te operationaliseren, zoals de Turing test [Copeland], en zijn de beroemde “robotica wetten” geformuleerd [Asimov].

Na 1950

In de jaren 50/60 heeft McCarthy de term *kunstmatige intelligentie* voor het eerst in de academische wereld laten vallen en zijn er demonstraties gegeven van een reeks werkende *weak AI* programma's, zoals *the Logic Theorist* [BBC Open University], *the General Problem Solver* en een eerste schaakspel [Buchanan]. Minsky heeft in deze jaren bewezen dat het mogelijk is om ruimtelijke en logische problemen van kleine omvang op te lossen en om simpele Engelse zinnen te begrijpen [BBC Open University].

In de jaren 70 zijn *expert systems*, taalanalyse systemen [BBC Open University] en de eerste robots ontwikkeld op de universiteit van Edinburgh, waarna er na 1980 een flinke versnelling is gekomen. De eerste *AI* producten werden op de markt gezet, de eerste “denkmachines” en neurale netwerken ontwikkelt en heeft Minsky baanbrekend onderzoek verricht op het gebied van het brein, waarin het brein als een collectie van samenwerkende agents wordt voorgesteld [Buchanan].

In de golfloorlog [BBC Open University] is er veelvuldig gebruik gemaakt van *kunstmatige intelligentie*, mede dankzij vooruitgang op het gebied van *machine learning*, *case-based & uncertain reasoning*, *natural language understanding* en *translation* [Buchanan]. Op het gebied van de neurale netwerken is er in de jaren 90 een vertraging ontstaan, hoofdzakelijk vanwege de performance-eisen van deze technieken en de ontdekking dat neuronen in de hersenen toch meer lijken op analoge pulsen dan digitale [Stanford].

Uiteraard is bekend dat in het laatste decennium veel autonome agents ontwikkeld zijn, vooral voor op het internet (*information retrieval*), is men met *Deep Blue* erin geslaagd om een machine te ontwikkelen welke beter kan schaken dan een mens en zijn ruimtevluchten zonder *AI* niet meer mogelijk.

Strong AI?!?

Hierboven is een korte, zeer beknopte samenvatting gegeven van de belangrijkste ontwikkelingen in de *kunstmatige intelligentie*. Echter, dit zijn allemaal *weak AI* voorbeelden. Software of robots met *strong AI*, welke bewustzijn en gevoel hebben, echte emoties kunnen tonen en qua intelligentie minimaal gelijkwaardig zijn, zijn tot op heden nog niet ontwikkeld. Hiervoor zijn nog te veel vragen op te lossen en technieken te perfectioneren. Zo is er op psychologisch vlak nog geen uitsluitsel over de wijze waaruit gevoel bestaat, hoe dit in elkaar steekt en hoe dit te simuleren c.q. emuleren is.

Maar is dit in de toekomst dan wel mogelijk? Op deze vraag zal in de volgende sectie worden ingegaan.

Toekomst

Er is een grote discussie gaande over de vraag of het mogelijk is om, in de toekomst, *strong AI* te kunnen ontwikkelen. Deze discussie bevindt zich op het moment vooral op filosofisch vlak. Om een goed oordeel te kunnen vormen zal ik in deze sectie een aantal meningen van deskundigen voorbij laten komen en tot slot een eigen oordeel vellen.

Het is per definitie niet mogelijk!

Roger Penrose [Penrose] heeft twee redenen geformuleerd waarom het, in zijn belevingswereld, onmogelijk is om *strong AI* te creëren. Zo geeft hij aan dat computers nooit intelligent kunnen zijn, aangezien formele systemen niet toereikend zijn in het kwalificeren van stellingen. Daarnaast moet een intelligent “systeem” bewustzijn hebben, iets wat niet te modelleren en dus niet te ontwikkelen is. Penrose gelooft dat de hersenen speciale biologische structuren bevatten om bewustzijn te bewerkstelligen, om zodoende *quantum effecten* uit te dragen. Dit zou niet mogelijk zijn in *AI*.

Edmund Furse [Furse] voegt er nog eens een vijftal, verzamelde, argumenten bij. Zo zou het onmogelijk zijn voor software/hardware om emoties te voelen en bewustzijn hebben, elementen welke essentieel zijn voor *strong AI*. Verder zou een computer nooit de juiste context kunnen vinden in conversaties (stelling van John Searle) om de juiste interpretatie te snappen, zou een computer geen eigen wil kunnen hebben en, ten slotte, een theologische reden. Deze vrije wil zou niet mogelijk zijn omdat mensgemaakte objecten alleen een voorspelbaar gedrag kunnen uitvoeren, aangezien mensen alleen maar voorspelbaar gedrag kunnen ontwikkelen. De theologische reden heeft betrekking op de aanname dat God ons, als enige “soort”, geschapen heeft als zijn evenbeeld, waardoor niemand ons intelligentie pijn zouden kunnen evenaren.

Verder worden er ook vraagtekens gezet bij het fenomeen creativiteit. Zou een computer creatief kunnen zijn?

De, hierboven gegeven, tegenargumenten zijn grotendeels eenvoudig te begrijpen en, eventueel, te weerleggen. Het belangrijkste tegenargument is filosofischer van aard en is geformuleerd door John Searle. Dit wordt ook wel het *Chinese Room* [Searle] argument genoemd.

Chinese Room¹

Stel, een persoon, welke geen Chinees kent, in een kamer opgesloten met een stuk tekst in het Chinees en heeft geen enkel contact met mensen buiten deze kamer. De persoon krijgt nog een tweede stuk tekst en een set regels in het zijn eigen taal. Deze regels stellen de persoon in staat om de goede symbolen terug te sturen als respons op het tweede stuk tekst. De mensen buiten de kamer noemen het eerste stuk tekst het ‘verhaal’, het tweede stuk tekst de ‘vragen’, en de instructies het ‘programma’. De symbolen die de persoon ze terugstuurt noemen ze de ‘antwoorden op de vragen’.

¹ Grotendeels overgenomen van <http://www.lapp-top.leidenuniv.nl/index.php3?c=34>

De persoon in de kamer kijkt dus bij de vragen, zoekt vervolgens in zijn instructies op welke serie symbolen moet volgen op de serie symbolen die in de vraag staat, en schrijft die vervolgens op. Op een gegeven moment wordt de persoon hier zo goed in, en de mensen buiten de kamer zo goed in het maken van de instructies, dat de antwoorden van de persoon in de kamer niet meer te onderscheiden zijn van de antwoorden van iemand die Chinees als moedertaal heeft.

De persoon in de kamer werkt als een computer: Hij zet symbolen om in andere symbolen. Volgens *strong AI* zou de goed geprogrammeerde computer de verhalen moeten begrijpen en zou het programma een uitleg zijn voor de werking van het menselijk brein. Searle stelt dat dit helemaal niet het geval is: De persoon in de kamer geeft de goede output bij een bepaalde input, maar hij heeft geen idee wat de symbolen die hij binnenkrijgt en terugstuurt betekenen. Het programma is ook geen uitleg voor het menselijk brein: als de persoon in de kamer een verhaal en vragen in zijn eigen taal zou krijgen, zou hij daar in zijn eigen taal antwoord op kunnen geven, die niet te onderscheiden zijn van iemand anders die die taal als moedertaal heeft.

Het uitvoeren van het programma stelt de persoon niet in staat om het verhaal, etc. te begrijpen, dus het uitvoeren van een programma stelt een computer niet in staat om het verhaal te begrijpen.

Het is wel mogelijk!

Mark Humphrys [Humphrys] geeft aan dat het in theorie mogelijk zou moeten zijn om *strong AI* te ontwikkelen, maar dat dit in de praktijk niet haalbaar zou zijn. Hierbij steunt Humphrys op het *materialism* [Trout] gedachtegoed. Hij vindt de beargumentatie van Penrose, Searle en Bringsjord ongeloofwaardig. Penrose heeft gelijk als hij beweert dat een computer meer nodig heeft dan een formeel systeem om waarlijk intelligent te zijn. Maar, zoals Humphrys ook beweert, geven Penrose argumenten geen onmogelijkheden aan. Waarom zouden creativiteit, gevoel en bewustzijn niet ontwikkeld kunnen worden?

Humphrys argumenten worden gesterkt door Edmund Furse [Furse] welke naast het verzamelen van tegenargumenten deze ook weerlegd.

Zo stelt Furse dat het onderzoek naar de menselijke emotie, zoals in Birmingham, goede vooruitzichten biedt om, uiteindelijk, modellen op te stellen over het fenomeen emotie. Wanneer dit mogelijk is, en niemand kan bewijzen dat het onmogelijk is, staat er niets in de weg om computers emotie te geven.

Hoewel Penrose beweert dat computers nooit bewustzijn kunnen hebben, geeft Dennett aan in zijn boek "Consciousness Explained" dat er weldegelijk een goed beeld is van het fenomeen bewustzijn en dat wij uiteindelijk dit begrip compleet kunnen begrijpen (en dus nabouwen). Het argument van Searle dat een computer de context achter bepaalde tekst niet zou kunnen begrijpen is volgens Furse achterhaald.

Furse geeft verder aan dat Margaret Boden bewezen heeft dat er geen limiet zit op de complexiteit van software en computers waardoor het, in de toekomst, mogelijk wordt voor computers om eigen keuzes te maken en dus een eigen wil hebben.

Het theologische argument wat is gegeven is moeilijker te weerleggen. Natuurlijk kan je de gehele theorie verwerpen, maar dat is gemakkelijk. Wat je wel kunt beweren is dat deze theologische theorie Christelijk van aard is. Het boeddhisme zegt echter dat het mogelijk is dat zielen van mensen naar dieren kunnen reincarneren. Daarnaast is het wetenschappers gelukt om apen een taal te leren [Lifson]. Dit impliceert dat deze apen ook een ziel hebben, en dus intelligent zouden zijn (zij het in mindere mate).

Creativiteit is een breed begrip. Wanneer is iemand creatief? Wanneer hij/zij een slimme oplossing bedenkt of wanneer iemand een kunstwerk maakt? In dit tweede geval is er reeds software beschikbaar welke, on-the-fly tekeningen kan produceren (<http://www.kurzweilcyberart.com/KCATaaron/STAFsample>).

Chinese Room

Er is veel geschreven over het Chinese Room theorie. In de literatuur zijn voorstanders te vinden, maar ook mensen welke deze theorie geheel verwerpen en beargumenteren waarom deze theorie niet opgaat.

Jake Copeland [Copeland] en Scott Crocheron [Crocheron] beweren in het zogenaamde *Logical Reply* dat dit onjuist is. Zo weten we uit de theorie dat de persoon geen Chinees kan. Maar het systeem welke bestaat uit de persoon, de tekst, de vragen, instructies, input en output, etc wel! Searle beweert dat als de persoon geen Chinees kan, het hele systeem het ook niet kan. Dit is echter, logisch gezien fout.

Copeland duidt i.h.b. op de ondeugdelijkheid van Searle's redenering: "De symboolmanipulatie zal de persoon niet helpen de input te begrijpen en de symboolmanipulatie zal het grotere systeem waarvan de persoon deel uitmaakt niet helpen de input te begrijpen.

Daarnaast is dit een gedachte-experiment. Wie zegt ons dat Searle een correct experiment heeft geformuleerd en is dit nou wel een realistisch voorbeeld en zegt dit experiment wat over een vervangend *AI* programma? Zijn wij wel slim genoeg om te kunnen concluderen dat dit gedachte-experiment wel correct is?

Bob Murphy [Murphy] geeft, tenslotte, aan dat de *Chinese Room* niets toevoegt aan de discussie over *AI* en eigenlijk maar onzin is.

Conclusie

Ondanks enige scepsis, heb ik geen argumenten kunnen vinden waarin ondubbelzinnig werd bevestigd dat het creëren van *strong AI* niet mogelijk zou zijn. Met neurale netwerken, agents technologie en verdere ontwikkeling van de belangrijke *AI* gebieden zal het in de toekomst mogelijk zijn om software te creëren welke net zo intelligent is als wij.

Maar willen wij dat wel? Hoe zal onze wereld er dan uitzien? Wat voor complicaties kunnen er dan optreden? In de volgende sectie zal ik hier kort op ingaan.

Ethische & maatschappelijke aspecten van waarlijk intelligente robots

Edmund Furse [Furse] denkt dat *strong AI* robots in de toekomst zeker geen slaven zullen zijn van ons. Immers, wanneer de robots net zo intelligent en onafhankelijk zijn als mensen, is er geen reden om aan te nemen dat wij de robots zullen onderdrukken. Daarnaast vind hij dat robots niet alleswetend kunnen zijn. Ze kunnen wel sneller een antwoord vinden op vragen.

Zoals aangegeven is in de video van Discovery Channel, geeft Furse ook aan dat het zeer waarschijnlijk is dat wij mensen robots zullen gaan opvoeden. Deze robots kunnen immers leren, zodat de kennis uit ervaring moet ontstaan.

Hoe moet er omgegaan worden met de rechten van robots? Wanneer robots eenzelfde mate van autonomie en intelligentie hebben als mensen, dan moeten deze robots als gelijke behandeld worden. Ook in de rechtzaal.

Furse lijkt ook vooruit naar de sociale aspecten van deze intelligente robots. Hij denkt dat de robots eenzelfde diversiteit zullen kennen als wij. De een zal willen trouwen en de andere robot zal zich willen afzonderen. Zouden gemixte relaties tussen robots en mensen voor kunnen komen? Wie weet!

Aangezien robots veel sneller kunnen denken en reageren, zal de communicatie tussen twee robots veel sneller en effectiever plaats kunnen vinden.

Zouden robots kunnen liegen? Hierover zijn de meningen verdeeld. Asimov [Asimov] denkt dat robots dit niet kunnen, aangezien ze “perfect” zullen zijn. Furse is een hele andere mening toegedaan. Hij neemt aan dat, aangezien robots ook bewustzijn en emoties zullen kennen, ook kunnen liegen. Ook omdat deze robots zullen “opgroeien” tussen mensen.

Wanneer robots bewust zijn van het feit dat ze liegen, zouden ze dan ook een notie kunnen hebben van, bijvoorbeeld, het Christelijke geloof? Furse denkt van wel. Immers, robots zullen nieuwsgierig zijn naar het ontstaan van de aarde en de mensheid en allerlei bronnen raadplegen. Wanneer de robot uitkomt bij de bijbel kan deze gelezen worden en kan er een waardering over uitgesproken worden. Ook door robots.

Furse stelt verder nog een erg interessant aspect aan de orde. Zullen robots ook doodgaan? In de letterlijke zin waarschijnlijk niet, aangezien mechanische componenten vervangen kunnen worden. Wat misschien wel mogelijk is, is een zelfmoordpoging. Mensen plegen zelfmoord wanneer ze geen uitweg meer zien of wanneer ze geen plezier meer halen uit het leven (depressie). Ook robots kunnen in theorie een depressie beleven, of uitgesloten worden uit het sociale leven aangezien ze erg verouderd zijn (mechanisch, maar ook geestelijk gezien). De robots zijn namelijk al lerend aan karakter, morele wetten e.d. gekomen, welke compleet anders zullen zijn van de robots die 3 generaties later ontwikkeld zijn. Misschien is dit ook wel de reden waarom wij een lichaam hebben welke niet het eeuwige leven heeft. Zouden wij niet gek worden als wij over 200 jaar nog zouden bestaan? Mijn opa's en oma's zijn al niet meer te “redden” voor het informatietijdperk, laat staan voor nog grotere veranderingen.

Moeten we wel intelligente robots ontwikkelen? Is dat onze taak wel? Deze vragen zijn ontzettend moeilijk te beantwoorden, Furse is van mening dat wij die wel moeten doen. Wij zouden kunnen profiteren van zulke robots, en wie zijn wij dan om ander *sapient life* te vernietigen. We moeten er echter wel van bewust zijn dat zulke robots ons werk zullen overnemen. Immers, zij kunnen ons werk sneller en effectiever doen. Aan de andere kant levert dit ons weer meer vrije tijd op.

Ik denk zelf dat de ontwikkeling op een gegeven moment niet meer tegen te houden zal zijn. De kennis zal hoe dan ook verder ontwikkeld worden. Je zou kunnen claimen dat de *strong AI robots* de volgende stap in de evolutie zouden zijn.

Zouden machines de mogelijkheid moeten hebben om levensbedreigende beslissingen te nemen, zoals het laten uitbreken van een oorlog of het bepalen of een patiënt in een ziekenhuis moet blijven leven? Velen denken dat Asimov's *robotic laws*² gehandhaafd moeten blijven. Mensen moeten dan altijd de controle blijven houden over

² **Law Zero:** A robot may not injure humanity, or, through inaction, allow humanity to come to harm.

Law One: A robot may not injure a human being, or, through inaction, allow a human being to come to harm, unless this would violate a higher order law.

Law Two: A robot must obey orders given it by human beings, except where such orders would conflict with a higher order law.

Law Three: A robot must protect its own existence as long as such protection does not conflict with a higher order law.

de robots. Robots moeten dan ook op een dusdanige manier ontwikkeld worden waardoor ze zich altijd handhaven aan deze wetten. Door deze intrigerende vragen, lijkt het mij verplicht om, wanneer wij dusdanige intelligente robots kunnen ontwikkelen, ook het morele aspect moeten bevatten. Robots moeten dan de frameworks bezitten om deze keuzes te maken.

Een andere, interessante, opinie van Nick Bostrom [Bostrom] vertelt het volgende: hoe kunnen wij discussiëren over morele en ethische aspecten van *strong AI* robots, wanneer deze intelligenter zijn dan ons. Wie zegt dat onze morele frameworks het beste zijn? Misschien missen wij wel intelligentie om de juiste beslissingen te maken. Wat is juist?

Conclusie

In deze essay heb ik onderzoek gedaan naar de verwachting of het in de toekomst mogelijk is om *strong AI* te kunnen ontwikkelen. Na het positioneren van de begrippen, waarbij ik intelligentie heb gedefinieerd als: “*a general mental capability that involves the ability to reason, plan, solve problems, think abstractly, comprehend ideas and language, learn and are conscious*”, is er gekeken naar de ontwikkeling van *AI*. Hieruit bleek dat tot op heden nog geen *strong AI* is ontwikkeld, waarna ik een literatuuronderzoek gedaan heb naar de vraag of dit in de toekomst dan wel mogelijk zou kunnen zijn.

In de literatuur zijn er veel voor en tegen argumenten gevonden. Vooral het *Chinese Room* argument is beroemd en berucht in deze. Na in eerste instantie gedacht te hebben dat er per definitie geen *strong AI* mogelijk zou zijn, heb ik redeneringen gevonden waarin het *Chinese Room* argument wordt ontkracht. Zodoende moet ik concluderen dat:

“Het per definitie niet onmogelijk is om in de toekomst strong AI te kunnen ontwikkelen.”

Ik denk dat deze ontwikkeling ook niet te stoppen is, maar dat dit niet binnen de komende decennia zal resulteren in intelligente robots. Hoe deze ontwikkeling dan maatschappelijk en ethisch te sturen is, is niet erg duidelijk en biedt ruimte voor filosofische discussies. Film als “I, Robot” en “AI” behoren daarmee dan wel tot de *science* maar niet tot de *fiction*.

Literatuur

Algemene informatie

Roger Penrose, *The Emperor's New Mind, Shadows of the Mind*

<http://www.comp.glam.ac.uk/pages/staff/efurse/Theology-of-Robots/Arguments-for-Strong-AI.html>

<http://www.comp.glam.ac.uk/pages/staff/efurse/Theology-of-Robots/Arguments-Against.html>

http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI15.html

<http://computing.dcu.ie/~humphrys/newsci.html>

http://www.onelang.com/encyclopedia/index.php/Artificial_intelligence#Strong_artificial_intelligence

<http://www.aaai.org/AITopics/html/ethics.html>

Chinese room

http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI14.html

<http://www.rpi.edu/~brings/LOG+AI/lai/node5.html>

<http://www.scottcrocheron.com/aipaper.htm>

<http://computing.breinstorm.net/searle+argument+chinese+semantics+address/>

<http://www.anti-state.com/murphy/murphy14.html>

Literatuurverwijzingen

AIII <http://www.aaai.org/>

Asimov <http://www.asimovonline.com/>

BBC http://www.open2.net/nextbigthing/ai/story_so_far/story_so_far.htm

Beer Heart of Enterprise, John Wiley & Sons, Chichester, 1979.

Diagnosing the System for Organizations, John Wiley & Sons, Chichester, 1985.

Bostrom <http://www.nickbostrom.com/ethics/ai.html>

Buitenhuis Viable, Adaptive Software, Eindessay Systeemtheorie, Radboud Universiteit Nijmegen.

Buchanan <http://www.aaai.org/AITopics/bbhist.html>

Cohen, Jonathan. Human Robots in Myth and Science. NY: A.S.Barnes, 1967.

Consoli Aantekeningen colleges

Copeland http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI15.html

Crocheron <http://www.scottcrocheron.com/aipaper.htm>

Discovery <http://tlc.discovery.com/convergence/robosapiens/robosapiens.html>

FreeDictionary <http://www.FreeDictionary.org>

Furse <http://www.comp.glam.ac.uk/pages/staff/efurse/>

Humphrys <http://computing.dcu.ie/~humphrys/newsci.html>

Imperial College London http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Conclusion

Lifson <http://people.cs.uct.ac.za/~flifson/things/aiassign1.html>

McCarthy <http://www-formal.stanford.edu/jmc/whatisai/whatisai.html>

Murphy <http://www.anti-state.com/murphy/murphy14.html>

Penrose, R. *The Emperor's New Mind, Shadows of the Mind*

<http://plato.stanford.edu/entries/whitehead/>

Searle <http://ist-socrates.berkeley.edu/~jsearle/>

Minds, Brains, and Programs, 1980, Cambridge University

Stanford university <http://www-cse.stanford.edu/classes/sophomore-college/projects-00/neural-networks/History/history2.html>

Syracuse university <http://www.phy.syr.edu/courses/modules/MM/ai/ai.html>

Tuten <http://huizen.daxis.nl/~henkt/intelligence-definition.html>

Trout <http://www.artsci.wustl.edu/~philos/MindDict/materialism.html>

Vuik <http://www.vuikvertier.nl/wetenschap/neurale-netwerken.htm>

Whitehead <http://plato.stanford.edu/entries/russell/>

Wikipedia http://en.wikipedia.org/wiki/Artificial_Intelligence

Wiener An introduction to Cybernetics. Chapman & Hall, London, 1957.

Wooldridge, Jennings Intelligent Agents: Theory and Practice, gepubliceerd in Knowledge Engineering Review, 1995.