

Gepubliceerd in KIJK nr. 9/2015

Wat als...

...robots slimmer worden dan mensen?

We vinden de wereld waarin we leven volstrekt vanzelfsprekend. Maar vul één feit anders in en hij ziet er plots volkomen anders uit. In deze serie kleuren we elke aflevering zo'n 'wat als...'-scenario in.

Tekst: Hidde Tangerman

Het had de nodige weerstand opgeroepen, maar Chris moest toegeven dat hij een beetje voorbarig was geweest met zijn oordeel. Schone luchten, nooit meer files, betaalbare zorg... robots hadden de wereld onmiskenbaar verbeterd. Goed, hij had zijn baan als brugwachter moeten opgeven – dat ging nu allemaal computergestuurd – maar als omgeschoolde IT'er had hij er eigenlijk veel leuker werk voor teruggekregen. Zijn schoonfamilie zat op de maan, daar hoefde hij ook niet meer op bezoek. Zelfs politici leken verstandige beslissingen te nemen nu de supersimulator dagelijks zijn algoritmen op de wereldproblematiek losliet. Immigratie werd eindelijk goed geregeld. Als ze door de RoboMarine uit zee werden gevestigd werden de asielzoekers door Immigration Computable eerlijk verdeeld over de Europese landen. Wat hij wel miste was een stukje rijden op de motor. Dat was met de intrede van zelfrijdend verkeer verboden. Veel te gevaarlijk. Daar stond tegenover dat de muziek beter was dan ooit. Wie had ooit gedacht dat machines kunst konden maken?

Een wereld waarin robots slimmer zijn dan de mens, het zou zomaar kunnen. Ontwikkelingen op het gebied van kunstmatige of artificiële intelligentie (A.I.) gaan razendsnel. "AI wordt zichtbaar, hoorbaar en tastbaar," zegt Guszti Eiben, hoogleraar kunstmatige intelligentie aan de Vrije Universiteit van Amsterdam. Volgens Eiben staat het tijdperk waarin robots veel van onze taken gaan overnemen voor de deur.

AI wetenschappers schatten de kans dat computers rond het jaar 2075 even intelligent zullen zijn als mensen in op 90 procent (Eiben zelf gelooft daar overigens niet in). De meesten

denken dat het daarna nog zo'n dertig jaar zal duren voordat computers 'superintelligent' zijn, wat wil zeggen dat ze de menselijke intelligentie op alle gebieden overtreffen.

Dit omslagpunt wordt singulariteit genoemd. Er bestaat ook veel scepsis over, maar als de verwachtingen van de optimisten uitkomen zullen we rond de volgende eeuwwisseling in een wereld leven die (deels) wordt bestuurd door computergestuurde robots. Hoe ziet zo'n wereld eruit?

Butlet-bot

De scenario's zijn zo divers als de talloze sciencefictionromans en -films die erover zijn verschenen. Zelf zijn de wetenschappers optimistisch; ongeveer 60 procent van recent gepolste A.I. deskundigen gelooft dat het in een wereld van *human level machine intelligence* goed toeven is. Zo'n 15 procent denkt dat de uitkomst min of meer neutraal zal zijn, terwijl een kwart een kwaadaardig of ronduit rampzalig scenario vreest.

In de positieve scenario's maken intelligente robots onze wereld gebruiksvriendelijker, vreedzamer, schoner en welvarender. Zo zal ons dagelijks leven een stuk aangenamer worden. Huishoudelijke taken zoals wasjes draaien, opruimen, en vuilnis buiten zetten besteden we uit de butler-bot. Onze robo-secretaresse plant onze agenda zo efficiënt mogelijk in, zodat onze dag altijd perfect gestroomlijnd verloopt. Maar intelligente robots kunnen meer dan alleen maar rekenen, poetsen of plannen. Met A.I. kan ook kunst worden gemaakt, zegt Eiben. "Een collega van mij uit Londen maakte met zijn evolutionaire algoritmen hij hele dansbare nummers. Daarmee was hij korte tijd heel populair in de *club scene*."

Nooit meer files

Ook rampenbestrijding profiteert van AI. Er zijn nu al robots die na een aardbeving mensen onder het puin kunnen opsporen. Kernrampen zullen in de toekomst tot minder schade leiden, aangezien robots – immuun voor radioactieve straling – direct na de ramp op locatie zijn om herstelwerkzaamheden uit te voeren.

Het is de vraag of rampen en storingen überhaupt nog plaatsvinden met alsmaar slimmere AI aan de knoppen. Zelfrijdende auto's zullen het aantal verkeersongelukken drastisch verlagen. Tegelijkertijd lossen ze ook het fileprobleem op, aangezien de verkeersstromen door de computer perfect op elkaar worden afgestemd. Intelligente robots brengen ook de kosten in de

gezondheidszorg flink omlaag. Met een geüploade database van de complete medische literatuur kunnen robots verbanden leggen die de huisarts over het hoofd ziet. Betere diagnoses betekenen dat nodeloze ingrepen worden voorkomen. “Deze technologie ligt nu al binnen handbereik,” zegt Eiben.

Zorgrobots kunnen het werk van verplegers en misschien zelfs chirurgen overnemen. Dat laatste roept interessante juridische vragen op. Wie draait er voor de schade op als er iets misgaat? De dienstdoende IT’er, de programmeur, de fabriek die de robot leverde, of misschien wel de robot zelf? Als robots intelligent worden kan het zijn dat ze ook moreel en juridisch aansprakelijk worden gesteld.

Aangezien alles waar reken- en papierwerk bij komt kijken, aan robots kan worden uitbesteed, zal er een hoop werk - denk aan dat van accountants, verzekeraars, secretaresses en baliemedewerkers - verdwijnen. Ook taxichauffeurs en vrachtwagenchauffeurs zijn niet langer nodig, omdat de logistieke sector zal worden geautomatiseerd. Toch hoeven we volgens John-Jules Meyer, hoogleraar kunstmatige intelligentie aan de universiteit van Utrecht, niet te vrezen voor grootschalige werkloosheid. “Deze vraag speelde al bij de industriële revolutie. Maar voor automatisering is altijd ander werk teruggekomen, dat vaak interessanter was. Als je robots inzet heb je IT’ers daaromheen nodig. Omdat de AI zich altijd doorontwikkelt, zullen er altijd nieuwe banen zijn.”

Net als echte evolutie

Kunnen robots ons ook helpen met de wereldproblematiek? Guszti Eiben: “Je kunt je voorstellen dat je een machine bouwt die de hele aarde kan vangen in één groot simulatiemodel, inclusief de economische activiteit, de groei van de bevolking, de opwarming van de aarde, enzovoort. Die simulator kan de regering beslissingen voorleggen.” Toch zal de mens nog altijd de knopen doorhakken, denkt Eiben. “We zullen zo’n systeem hooguit als adviseur willen, die alleen een scenario-analyse maakt. Maar wij brengen het in de praktijk.” De van origine Hongaarse hoogleraar is zelf gespecialiseerd in het ontwikkelen van AI door evolutie. Daarbij worden machines zo geprogrammeerd dat ze van hun ervaringen leren en hun gedrag voortdurend aan de omstandigheden aanpassen, net zoals bij ‘echte’ evolutie. Momenteel werkt Eiben aan een baanbrekend onderzoek om robots zichzelf te laten kopiëren. Deze *self-reproducing machines* kunnen worden ingezet op plekken waar de mens moeilijk

kan komen. In de toekomst gaan we bijvoorbeeld niet zelf de mijnen in, maar sturen we de robots samen met een 3D-printer naar beneden. Via evolutie en voortplanting ontstaat zo een kolonie robo-mijnbouwers die de grondstoffen efficiënter en veiliger delft dan de mens ooit zou kunnen.

Een andere mogelijke toepassing die Eiben noemt, is ruimtekolonisatie. “Omdat je de omstandigheden op een nieuwe planeet niet kent, kun je je robots er niet voor optimaliseren. Wel kun je ze uitrusten met de mogelijkheid om te evolueren. Je stuurt een kolonie die zich afhankelijk van de omgeving ontwikkelt en de hele planeet voor ons geschikt maakt. Het is vergezocht, maar wel denkbaar.” In het intelligente machinetijdperk, waarin deze techniek is vervolmaakt, zullen we dus onze eigen woonplaats elders in het zonnestelsel kunnen uitkiezen. Maan, Mars, Titan, *here we come!*

Aangezien evolutie van nature onvoorspelbaar is, zou het ook mis kunnen gaan. De robots zouden een kant op kunnen evolueren die minder prettig voor ons is. Een zelfrijdende auto kan bijvoorbeeld veranderen in een moordmachine. Eiben: “Daarom moet je de AI zo bouwen dat er harde grenzen zijn die niet overschreden kunnen worden. Er is altijd een ruimte van opties waarbinnen de robot keuzes maakt en dingen uitvoert. Hoe die ruimte eruitziet wordt bepaald door de ontwerper. Als we die ruimte opstellen met veilige opties, dan kan zo’n doemscenario niet plaatsvinden.”

Game over?

Iemand die wél in doemscenario’s gelooft, is Nick Bostrom, een Zweedse hoogleraar filosofie aan de universiteit van Oxford die zich bezighoudt met nadenken over computerintelligentie. Zijn recente boek *Superintelligentie* gaat over de mogelijke gevolgen van een wereld waarin machines superintelligent en zelfbewust worden.

Bostrom gaat ervan uit dat een superintelligent systeem – dat kan een robot zijn, maar net zo goed een stuk computersoftware – over ‘superkrachten’ beschikt waarmee het zo goed als almachtig kan worden. Via grenzeloze kennis, perfecte strategische planning en uitgekiende manipulatietechnieken kan een superintelligentie al snel de macht grijpen op aarde, betoogt de Zweed.

Wat een superintelligentie vervolgens gaat doen hangt af van zijn ‘einddoel’; de taakomschrijving die er door de menselijke ontwerpers is ingeprogrammeerd. Zelfs als dat

eindoel goedaardig is, kan dat leiden tot een rampzalige uitkomst, waarschuwt Bostrom. Een AI wiens taakomschrijving is om ons te laten glimlachen zal dat doel in de beginfase – als het nog relatief ‘dom’ is – nastreven door amusant gedrag te vertonen. Maar zodra hij superintelligent wordt, ontdekt hij dat het efficiënter is om elektrodes in ons brein te plaatsen die onze gezichtsspieren in een permanente grimas houden. Geen prettig vooruitzicht. Of stel dat de taakomschrijving het oplossen van wiskundige problemen is. Om maximaal resultaat te behalen, zou de superintelligentie onze gehele planeet – en zelfs het gehele universum – kunnen veranderen in één grote computer. Niet echt wat we bedoelden, maar volstrekt logisch voor de AI. “Veel einddoelen die op het eerste gezicht veilig en verstandig lijken, blijken bij nader inzien radicaal onbedoelde consequenties te hebben,” schrijft Bostrom. “Als een superintelligentie met één van deze einddoelen een beslissend strategisch voordeel behaalt, is het *game over* voor de mensheid.”

De uitdaging is dan ook om de AI een taakomschrijving mee te geven die overeenkomt met onze eigen waarden en normen. Bostroms suggereert om dit aan de intelligentie van de machines zelf over te laten, bijvoorbeeld door de doelstelling te formuleren als: “Bereik wat wij zouden willen dat de AI zou bereiken als we er lang en diep over zouden hebben nagedacht.”

Stephen Hawkings angst

De Nederlandse hoogleraren hebben het boek van Bostrom gelezen, maar delen zijn analyses en conclusies niet. Eiben: “Dat machines zelfbewustzijn verkrijgen is zéér onwaarschijnlijk, en dat het dan een doemscenario zou worden is ook zéér onwaarschijnlijk.”

Volgens Meyer berust het idee van singulariteit sowieso op een denkfout. “Om singulariteit te bewijzen wordt altijd de wet van Moore aangehaald, die stelt dat het rekenvermogen en de geheugencapaciteit van computers elke twee jaar verdubbelen. Dat geldt misschien voor de hardware, maar de software en de onderliggende theorie groeien helemaal niet zo hard. Juist op die gebieden moet je een doorbraak krijgen als je naar superintelligentie wil.”

Toch zijn er genoeg serieuze wetenschappers die heilig geloven in singulariteit. Zelfs Stephen Hawking waarschuwde onlangs nog voor kunstmatige intelligentie, die hij “potentieel gevaarlijker dan kernwapens” noemde. Meyer: “Ik heb Stephen Hawking heel hoog zitten,

maar ik begrijp de ratio niet voor zijn angst. Ik denk eerlijk gezegd dat hij er te weinig van weet. Juist door zijn handicap kan hij enorm van AI profiteren.”

KADER: Sterke of zwakke A.I.

De vraag wat er precies wordt bedoeld met (super)intelligente robots raakt aan een decennia oude splijtzwam in de wereld van kunstmatige intelligentie, namelijk de tweedeling tussen ‘sterke’ en ‘zwakke’ A.I. Wetenschappers uit het kamp van de sterke A.I. geloven in de mogelijkheid dat een computer of robot autonome intelligentie of zelfbewustzijn kan ontwikkelen en dus net zo zal kunnen functioneren als de mens zelf. Aanhangers van de zwakke A.I. theorie zeggen dat een computer misschien intelligent of autonoom lijkt, maar dat niet werkelijk is. Hij moet altijd binnen de lijntjes kleuren die wij voor ze hebben getrokken. Er is vooralsnog geen wetenschappelijk bewijs gevonden voor één van beide stellingen. Scenario’s van een wereld geregeerd door robots zullen dus verschillen al naar gelang je in sterke of zwakke A.I. gelooft.

Abhinabha Tangerman sprak met voor dit artikel met prof. Dr. Gusztai Eiben, hoogleraar Computational Intelligence aan de Vrije Universiteit van Amsterdam en met prof. Dr. John-Jules Meyer, hoogleraar Informatica en Intelligente Systemen aan de universiteit van Utrecht en wetenschappelijk directeur van het Alan Turing Institute Almere. Hij raadpleegde de volgende literatuur:

- Nick Bostrom: *Superintelligentie* | Uitgeverij De Wereld (2015)
- Bard van de Weijer: *Evoluтиetheoriepraktijk* | De Volkskrant (19 april 2014)