

# Kies de juiste voorbeelden!

Kunstmatige intelligentie (Artificial Intelligence of kortweg AI) verovert de wereld in hoog tempo. In ons dagelijks leven weten we soms niet meer of we bij een helpdesk een mens van vlees en bloed aan de telefoon hebben of een AI-gestuurde voice- of chat-bot. Die veranderingen hebben zeker gevolgen voor de business - of informatieanalist. Maar welke? Moeten we andere vaardigheden aanleren? Om ons hier een beeld van te vormen gaan we te rade bij Arjan van Hessen, onderzoeker taal- en spraaktechnologie.

door Reinoud de Leve en Hans Siebering



Bij het afnemen van dit interview hebben we gebruik gemaakt van AI. De dag na het interview kregen we van Arjan de vertaling door AI van onze opname van het gesprek naar geschreven tekst. Die bestond uit 20.000 woorden. Zes keer zoveel als de uitwerking van het interview hieronder. De machine pikt opmerkelijk veel, soms gemompelde details op van wat er gezegd wordt. Ook het scheiden van de verschillende sprekers gaat meestal goed. Het valt op dat mensen een gesprek niet in zinnen voeren, maar in stukjes van zinnen. Dat leest niet fijn. Daarom is het vervolg van dit artikel het door ons geredigeerde interview.

## **Wat is AI?**

AI is een containerbegrip, maar we kunnen het omschrijven als 'de studie van hoe computers te trainen, zodat ze dingen kunnen doen die wij mensen op dit moment beter kunnen doen'. Het is een poging om menselijke capaciteiten toe te voegen aan een computer.

De manier waarop dat tegenwoordig veel gebeurt, is via *Machine Learning* (ML). De computer krijgt heel veel voorbeelden aangeboden die gekoppeld zijn aan een juist antwoord of een bepaalde actie. ML is dus een computerapplicatie die gebruikt wordt om computers te laten leren van data en daarmee soms een vorm van menselijke intelligentie te geven. Het werkt dus niet, zoals in het verleden met vooraf ingebrachte regels: als de temperatuur hoger is dan zoveel graden, druk dan dat knopje in. Die manier van werken is nog steeds nuttig, maar het is echt wat anders dan ML. Wat je met ML zou kunnen doen is dat je bij een populatie van mensen registreert wanneer ze op dat knopje drukken.

# Arjan van Hessen over AI

Het resultaat stop je in een zelflerend algoritme. Dat schat dan in dat als de temperatuur op vrijdag boven de 37 graden komt, mensen op dat knopje drukken. Op donderdag gebeurt het – niemand weet waarom – al bij 36,5 graden en op zondag pas boven de 40 graden. De informatie zit in de verzamelde data en die proberen we met ML eruit te halen. De regels worden dus niet meer door mensen ingebracht, maar de machine leidt de regels zelf af uit ons gedrag. De machine kan dat dikwijls veel beter en sneller dan mensen. Maar de machine kan geen verklaring geven. Als de verklaring is dat het komt doordat de poriën van mensen op zondag verder open staan dan hebben we nog steeds mensen nodig om dat te bedenken. Het resultaat is nu een computerapplicatie die op het juiste moment op het knopje drukt: we noemen dat dan [Kunstmatige Intelligentie](#).

De regels worden niet meer door mensen ingebracht, maar de machine leidt de regels zelf af uit ons gedrag. De machine kan dat dikwijls veel beter en sneller dan mensen. Maar de machine kan geen verklaring geven.

## Hoe gaat het leren in zijn werk?

Voordat AI in gebruik genomen wordt kies je eerst een grote dataverzameling, die je als leerstof gebruikt. Die verdeel je in drie gelijke delen.

- Het eerste deel gebruik je om de AI te *trainen*: op deze foto staat een kraai, op die een paard, etc.
- Het tweede deel gebruik je om te *tunen*: je vraagt 'wat is dit?' en verbetert de antwoorden.
- Met het derde deel van de data ga je *testen* en vaststellen hoe goed de AI het doet.

Vervolgens verdeel je de oorspronkelijke dataverzameling op een andere manier in drie delen en voer je dezelfde drie stappen uit. Nadat je dit ongeveer tien keer hebt herhaald mag je ervan uitgaan dat het gemiddelde van de tien testen een redelijke voorspelling is van wat er gebeurt als je de AI op nieuwe data toepast. Je hebt hiervoor veel data nodig van hoge kwaliteit. Gemiddeld zit zestig procent van het werk in het schoon maken van de data. Bovendien zul je het algoritme altijd bij moeten blijven trainen omdat mensen en de wereld om ons heen veranderen.

Machine learning, de manier waarop de computer kan leren, kan op verschillende manieren gedaan worden. De enorme vlucht die AI de laatste jaren heeft genomen, komt doordat ML gedaan wordt met behulp van zogeheten *Neurale Netwerken* (NN). NN bestaan al decennia maar de enorm toegenomen rekenkracht en de beschikbaarheid van heel veel data, maakt het tegenwoordig mogelijk om Neurale Netwerken met

Voor het trainen, tunen en testen heb je veel data nodig van hoge kwaliteit. Gemiddeld zit zestig procent van het werk in het schoon maken van de data.

verschillende lagen te maken: de *Deep Neural Networks* (DNN). Het gebruik van al die lagen is de magische truc, waardoor het nu zo ontzettend goed gaat. Een netwerk heeft bijvoorbeeld 32 of 64 lagen. Aan de eerste laag bied je een beeld of geluid aan. Met behulp van wat de AI heeft geleerd en een hoop hogere wiskunde wordt aan het patroon laag na laag steeds meer betekenis gegeven. De laatste laag zegt: op deze foto staat een 'kraai' of 'paard'. Met regelgebaseerde systemen lukt dit niet. Probeer maar eens te beschrijven waarom het object op de foto een paard is (en kom dan niet aan met: het lijkt op een paard).

## Hoe 'menselijk' is AI op dit moment?

Er zijn nog wel een paar punten waarin AI duidelijk verschilt van menselijke intelligentie. AI correleert input en output: 'Als de haan kraait, dan gaat de zon op'. Mensen proberen oorzakelijke verbanden te leggen: 'De haan kraait *omdat* de zon opkomt'. Maar dat laat onverlet dat AI steeds meer menselijke trekjes krijgt. Het duidelijkst zien we dat op het vlak dat ons van alle andere dieren onderscheidt: taal. Natuurlijk kennen dieren ook vormen van taal: het optrekken van de lippen van een wolf, het kwispelen van een hond, het zoemen van een bij. Maar dit zijn basale taaluitingen: kom niet dichterbij want dan bijt ik! En ook zal je hond je niet vertellen dat ze vandaag niet maar morgen wel naar dat bos wil om te gaan ravotten.

Iets dergelijks zien we bij emotie. Mensen kennen net als zoogdieren basale emoties: de reactie van onze [hersenen](#) op een [affectieve](#) prikkel (vreugde, angst, boosheid, verdriet). Maar wij mensen kennen ook het gevoel dat gezien kan worden als een speciale uiting of vorm van emoties: de bewuste beleving, of mentale reflectie van een emotie (sarcasme, ironie, humor). AI kan steeds beter die basale emoties detecteren, maar met die andere vormen van emotie lukt het nog niet echt. Dat komt onder andere doordat die 'mentale reflecties emotie' enorm context gevoelig zijn. Iets dat wij leuk vinden, kunnen Italianen als een belediging opvatten. Een opmerking van een Engelsman ('a very interesting point of view') wordt door ons Nederlanders dikwijls als een compliment gezien terwijl het zo niet bedoeld is.

Het zal duidelijk zijn dat de context van talige uitingen enorm belangrijk is en dat AI, wil het een plek tussen ons mensen veroveren, daar rekening mee zal moeten houden. Een pratende zorgrobot in Frankrijk zal duidelijk verschillen van de Nederlandse 'collega' omdat de

Het zal duidelijk zijn dat de context van talige uitingen enorm belangrijk is en dat AI, wil het een plek tussen ons mensen veroveren, daar rekening mee zal moeten houden.

Franse sociale conventies nu eenmaal anders zijn dan die in Nederland. En dat levert interessante problemen op. De privacy van mensen in de VS is al niet geweldig en in China is die volledig afwezig. Om AI te trainen is veel data nodig en dus is het veel makkelijker om de 'menselijke taal' (gesproken en geschreven) in Amerika en China te verzamelen: daar mag ongeveer alles. Maar het gevolg is wel dat talige-AI het risico loopt een Amerikaanse of Chinese *bias* te krijgen, en daar zitten we niet op te wachten. AI in Nederland zal zich toch moeten voegen naar de Nederlandse gebruiken en niet eisen dat wij ons voegen naar de manier waarop er bv. in Amerika wordt gecommuniceerd.

## Hebben wij invloed op wat de systemen leren?

Zoals gezegd, leert AI van voorbeelden: op deze foto staat een 'kraai', op deze een 'paard', deze gesproken uiting schrijf je als 'appel', deze als 'appèl'. Zolang wij er maar de juiste data instoppen, zal het leren wel goed gaan. Maar het verzamelen van 'goede' data is een enorm tijdsintensief karwei. Dus wat is er makkelijker dan gewoon het internet opgaan en daar data van ons mensen ophalen om te trainen.

Google heeft dat gedaan voor het herkennen van poezen op foto's. Als je foto's ophaalt waarvan het bijschrift het woord 'poes', 'kat' of 'kater' bevat, dan mag je ervan uitgaan dat er een poes op de foto zal staan. Dat werkte goed, maar het ging fout toen ze probeerden apen te herkennen. Veel foto's van donkergekleurde mensen bevatte bijschriften van het type 'het zijn net apen' en dus kun je foto's van donkergekleurde mensen krijgen als je zoekt op 'apen'. Dat was dus niet omdat Google bewust discrimineerde, maar wel omdat wij mensen dat soort foto's en teksten op het internet zetten. Hier geldt heel duidelijk: *garbage in, garbage out*.

Dit voorbeeld is duidelijk en bijna niemand zal betwisten dat het resultaat ongewenst is. Maar het trainen van AI stelt de samenleving wel voor morele dilemma's. Agenten die de rapper Typhoon aanhielden, hebben dat gedaan op basis van de combinatie dure auto, jonge leeftijd en donkere huidskleur. Dat is etnische profilering die gebaseerd kan zijn op al dan niet correcte ervaringsgegevens. Maar als ik vijf keer door een oudere dame met grijs haar in elkaar ben gemept, dan ga ik een straatje om wanneer ik zo iemand in een donker steegje tegenkom. Als mensen dit soort gevolgtrekkingen maken, lachen we erom of halen we onze schouders erover op, maar als software hetzelfde doet staat iedereen op zijn achterste benen: discriminatie!

AI moet ook dit soort beslissingen nemen maar met welke data gaan we trainen? Als je wilt dat AI alleen maar politiek correcte beslissingen neemt, dan moet je ervoor zorgen dat de voorbeelden waarmee getraind gaat worden ook politiek correct zijn. Maar wie gaat bepalen wat correct is? Dit is een gevoelig onderwerp. We verwachten van software dat die onpartijdig en efficiënt is. Maar wat als de data een ander patroon laten zien dan we wenselijk vinden? Oplosbaar is dit dilemma waarschijnlijk niet, en daarom moeten we bij het trainen en inzetten van AI ons bewust zijn van dit soort mogelijke *biassen*.

We verwachten van software dat die onpartijdig en efficiënt is. Maar wat als de data een ander patroon laten zien dan we wenselijk vinden?

## Regelgebaseerde systemen bevatten kennis. Een regelgebaseerd systeem kan in principe laten zien hoe het tot een beslissing is gekomen. Hoe doen neurale netwerken dat?

Neurale netwerken kunnen dat niet automatisch. Het is een actueel onderwerp. De EU heeft onlangs aangekondigd dat het een paar miljard vrij gaat maken voor onderzoek, waarbij één van de speerpunten *explainable AI* (XAI) is. XAI is een nieuwe richting in dit vak: leg uit waarom je die beslissing hebt genomen. Er wordt niet meer geaccepteerd dat AI zegt: 'sorry, u krijgt geen hypotheek'. XAI moet dan laten zien op grond van welke informatie die beslissing werd genomen. Het feit dat je in een wijk wilt gaan wonen met meer dan gemiddeld slecht betalende bewoners, zal dan waarschijnlijk niet geaccepteerd worden als motivatie om de hypotheek af te wijzen.

## Wie is verantwoordelijk voor de adviezen en beslissingen van AI?

Ook dat is een ethische vraag. Het antwoord hangt af van de toepassing. De waarden van de samenleving waarin de AI getraind is, worden weerspiegeld in de beslissingen die de AI neemt. Naarmate AI meer ingezet wordt, spelen ethische vragen een grotere rol. Daarom móét de samenleving daarover nadenken. De uitkomsten van die discussie moet je in de trainingsdata verwerken. Als je wilt dat de AI niet op leeftijd of gender discrimineert (hier: onderscheid maakt) kun je het beste leeftijd en gender niet in je trainingsdata opnemen.

Met AI hebben we een zeer krachtige technologie in handen die, dankzij de beschikbaarheid van grote hoeveelheden data, in staat is om veel en uiteenlopende taken te leren. Maar het heeft er ook voor gezorgd dat AI-systemen ongewenst (racistisch of seksistisch) gedrag kunnen vertonen. We moeten ons daarom goed

# Arjan van Hessen over AI

afvragen hoe en waarvoor we de technologie willen inzetten. En in welke mate AI-gedreven systemen zelfstandig beslissingen mogen nemen. Daar ligt een gezamenlijke taak voor AI-experts en ethici.

## **Hoe wordt AI momenteel toegepast?**

AI wordt veel en veel meer ingezet dan de meeste mensen vermoeden. Het wordt gebruikt in allerlei kennissystemen, in systemen die de stoplichten regelen, in het beslissen waar wanneer de ambulances moeten staan voor een optimale dienstverlening of wanneer en waar er gespreid moet worden en meer.



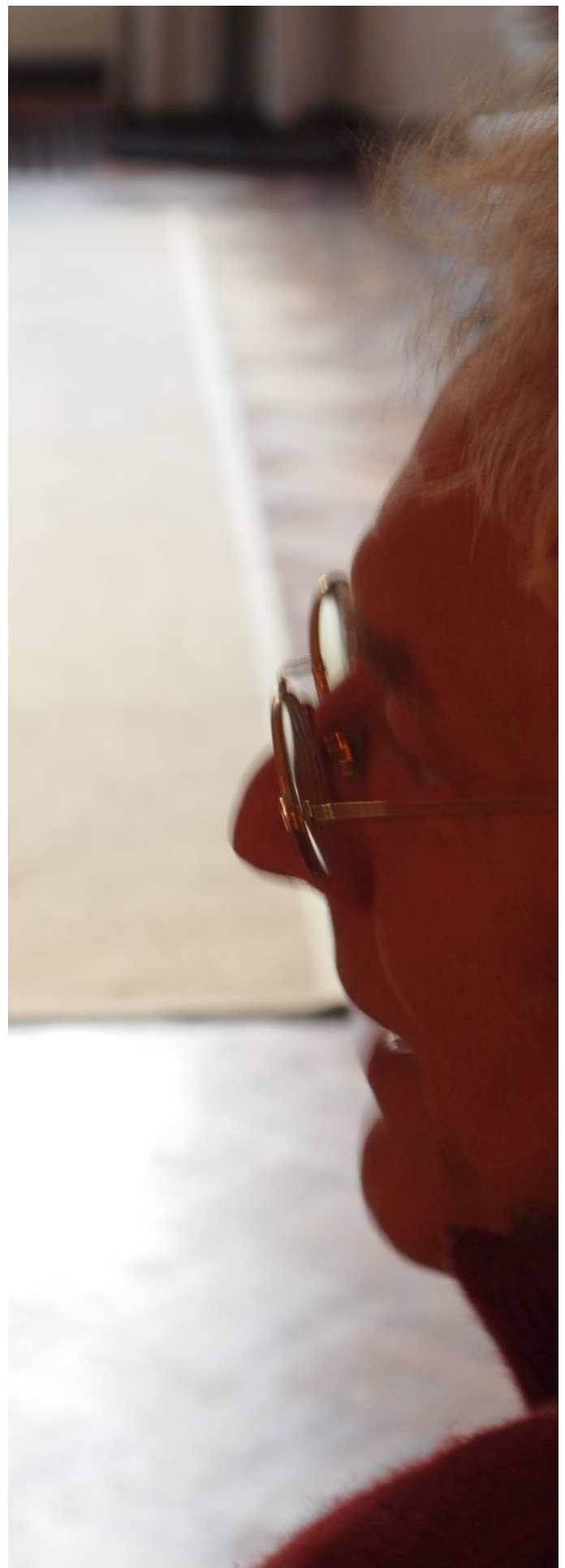
Een spannend onderdeel van de huidige AI, is de poging om van data naar begrip te gaan. Als ik op Google zoek naar 'hoe ontstaat regen' dan krijg ik nu een lijst van 1,4 miljoen artikelen waarin mogelijk het juiste antwoord staat. Maar die wil ik niet allemaal lezen: ik wil een antwoord. Met behulp van het zogeheten semantisch web wordt geprobeerd om kennis te extraheren uit de beschikbare data. Google geeft zelf al aan hoe ze dit ongeveer doen.



Het idee achter het semantisch web is een beetje als volgt: probeer de kennis op te slaan middels zogeheten *triples*. Triples bestaan uit drie onderdelen: een subject, een predicaat en een object. Als voorbeeld kun je denken aan:

1. 'Amsterdam' -> 'is hoofdstad van' -> 'Nederland'
2. 'Amsterdam' -> 'heeft aantal inwoners' -> '821K'
3. 'Amsterdam' -> 'gesticht in' -> '1275'

Dit zijn drie heel banale feitjes. Maar stel dat we dat voor alle plaatsen op de wereld doen, dan kunnen we vragen beantwoorden als 'welke hoofdsteden ouder dan 400 jaar hebben minder dan 1M inwoners?' Door enorm



# Interview

grote hoeveelheden van dit soort gegevens op te slaan en met elkaar te verbinden, kunnen we kennis van de wereld opslaan en ontsluiten. Dit staat ook wel bekend als *Linked (Open) Data*. Talige AI-software die 'weet' wat het onderwerp en gezegde in een zin is, kan met behulp van de opgeslagen triples antwoord geven op vragen. Het bekendste voorbeeld is IBM's Watson dat er in 2011 in slaagde het quiz-programma *Jeopardy* te winnen. Watson had haar kennis vergaard door het lezen van enorme hoeveelheden Engelstalige kranten, boeken en websites en de daar opgeslagen informatie om te zetten in triples en daarover te 'redeneren'.

Het idee achter het semantisch web is een beetje als volgt: probeer de kennis op te slaan middels zogeheten triples. Triples bestaan uit drie onderdelen: een subject, een predicaat en een object.

In de bedrijfsvoering zien we op dit moment ontwikkelingen waarbij AI niet de eindbeslissing neemt, maar een advies geeft aan een expert, die vervolgens de beslissing neemt. De AI is dus ondersteunend en daar heeft AI momenteel de grootste meerwaarde. Stel dat de AI zegt: deze patiënt is uitbehandeld, maar de arts weet dat hij binnenkort een kleinkind krijgt en dat nog graag mee wil maken. Dan kan de arts beslissen om toch nog chemo te geven. Ik vind het een prettig idee dat de mens in dit soort situaties in control blijft. Deze zogenoemde *assisting technology* is dan ook ethisch veel zuiverder.

## Gaat er een singulariteit komen?

Singulariteit is een ambigu begrip, want de voorspellers daarvan zien kennis en rekenkracht als hetzelfde. Het idee is dat de chips alsmat sneller worden (Wet van Moore), meer data kunnen verwerken en een grotere reken capaciteit krijgen. De mens heeft bijna 90 miljard neuronen. Je kunt voorspellen wanneer de rekenkracht van een chip die van de mens zal overtreffen. Dat zou rond 2025 kunnen zijn. Maar volgens mij moet je niet alleen naar de reken capaciteit kijken, maar vooral naar de begrip capaciteit. Een Afrikaanse Olifant heeft 3x meer neuronen dan een mens maar is *toch minder intelligent*. Om het moment te bepalen waarop de chip de mens daarin overtreft hebben we een goede operationele definitie van 'begrip' nodig. Helaas is die er nog niet.

Maar dat er een moment van singulariteit gaat komen, daarover zijn de geleerden het wel eens. Het is alleen de vraag wanneer. De één zegt over dertig jaar, de ander over honderdvijftig jaar. Maar wanneer het komt moeten we weten wat we ermee willen en hoe we er mee willen omgaan. We moeten ons er daarom goed op voorbereiden.

De **Singulariteit** is de hypothetische toekomstige creatie van superintelligente machines. Superintelligentie wordt dan gedefinieerd als een technologisch gecreëerde cognitieve capaciteit die veel verder gaat dan mogelijk is voor mensen.

Mocht de Singulariteit zich voordoen, dan zal de technologie verder gaan dan ons menselijk vermogen om de resultaten ervan te voorzien of te beheersen en zal de wereld onherkenbaar worden getransformeerd door de toepassing van superintelligentie op mensen en/of menselijke problemen, waaronder armoede, ziekte en sterfte.

Neem bijvoorbeeld de Crispr-Cas technologie. Opeens was het er en nu wordt het toegepast, terwijl niet goed nagedacht is over de gevolgen en risico's. Als de singulariteit eraan komt lukt het echt niet om binnen bijvoorbeeld een halfjaar wereldwijd te bepalen wat we willen. Je bent heel lang bezig met uitwerken en testen van scenario's om vast te stellen wat voor maatschappij we willen waarin AI zo'n dominante rol speelt.

## Het punt bij singulariteit is niet alleen rekenkracht en begrip. Sommigen voorspellen dat AI de macht gaat grijpen. Is dat een realistische verwachting?

Ja en nee. In China zijn ze nu bezig met een sociaal kredietsysteem, dat de gezichten van 1,3 miljard Chinezen kan herkennen, en ook, tot een afstand van 50 meter, mensen aan hun lichaamshouding. Het systeem weet wie er praat en wat er gezegd wordt. AI die dingen worden, samen met wat andere mensen van jou vinden, gebruikt om een waardeoordeel over jou te geven: een sociaal krediet. Het is een puntensysteem, op basis waarvan bepaald wordt welke vrijheden je krijgt. Als je onder een bepaald niveau komt, wordt je hypotheek duurder, mag je niet naar het buitenland of mogen je kinderen niet meer naar de universiteit. Het systeem is nu nog op vrijwillige basis en wordt rond 2020 verplicht. Voor zover ik weet is er geen beroep mogelijk tegen door het systeem gegeven ratings. Op het moment dat je zo'n systeem die macht geeft en hooguit met steekproeven controleert of de beslissingen nog reëel zijn, kom je wel in de gevarenzone. De volgende stap zou kunnen zijn: richt het systeem zo af dat het economisch zo goed mogelijk gaat of dat de kans op gele hesjes in China zo klein mogelijk is. Dan gaat het zelf de grenzen beïnvloeden. Dan gaat het in de richting van systemen die een grote mate van zelfstandigheid hebben in het bepalen van het leven van mensen. Een voorbeeld van hoe dat eruit kan gaan zien kun je zien in de aflevering *Nosedive* van de serie *Black Mirror* op *Netflix*.

# Arjan van Hessen over AI

Het is de vraag of de ontwikkeling in de richting van dergelijke systemen wel tegen te houden is. Zijn we in staat om nee te zeggen tegen iets dat 'in de praktijk best goed werkt'? Ga jij nog naar een echte dokter als de supercomputer Watson van IBM het, zonder wachttijd net zo goed doet voor tien procent van de prijs? Wij moeten als maatschappij in staat zijn om nee te zeggen tegen alleen dat economisch voordeel van autonoom werkende AI-systemen. Of dat lukt vraag ik me af.

***In plaats van de vraag te beantwoorden of machines konden denken, stelde Alan Turing een experiment voor waarin mensen moesten beslissen of het antwoord op een gestelde vraag van een mens of een computer afkomstig was. Kan een machine een mens zodanig imiteren, dat een mens het verschil tussen mens en machine niet meer ziet? Door dit experiment hoefde Turing geen antwoord te geven op de vraag wat 'menselijk' denken feitelijk inhield. Inmiddels hebben veel machines deze test min of meer doorstaan. Is het nu misschien***



**Arjan van Hessen** is afgestudeerd in geofysica en Italiaans en gepromoveerd in fonetiek. Hij heeft gewerkt bij het Belgische bedrijf Lernout en Hauspie, dat een pionier is op het gebied van spraaksynthese.

Tegenwoordig is hij werkzaam als onderzoeker taal- en spraaktechnologie bij de universiteiten van Twente en Utrecht en het Enschedese bedrijf Telecats.

***wel zaak om die vraag waarin 'menselijk' denken zich onderscheidt van 'artificieel' denken te beantwoorden?***

Het is niet waar dat er machines zijn die de Turing test hebben doorstaan, tenzij je naar heel specifieke deelgebieden kijkt. Je kunt AI maken, die heel goed een taart kan bakken of een afspraak met de kapper kan maken. Maar dat wil niet zeggen dat zo'n systeem antwoord kan geven op de vraag of hij meer houdt van de zomer of van de winter. AI, die een mens in al zijn facetten kan imiteren, is er nog lang niet. Jij kunt bakken, autorijden, leren sokken te breien en je kunt antwoord geven op de vraag wie daar loopt. Er is geen software op de wereld die dat allemaal kan. Voorlopig nog niet althans.

***Toen we in ons vak voor bepaalde problemen regelgebaseerde oplossingen gingen maken, veranderde dat het werk van de analist. Daarvoor was de analist gewend om een hele beslissing uit te tekenen in blokjes en wybertjes. Nu hoefde hij dat niet meer te doen, maar moest hij een set 'kennisregels' definiëren. Welke vaardigheden heeft een analist voor de huidige en toekomstige AI-systemen nodig?***

AI kan door heel veel voorbeelden aan te bieden getraind worden om complexe bewerkingen uit te voeren of complexe beslissingen te nemen. Of die nu in alle omgevingen, zoals de Belastingdienst gebruikt mogen worden weet ik niet, want daar zit een duidelijk juridische component aan. Maar waar AI toegepast mag worden zal de rol van de analist meer gaan liggen in enerzijds het bedenken van de juiste voorbeelden waarmee de machine getraind mag worden en anderzijds het vertalen van de uitkomsten van AI naar de praktijk van alledag.

Maar... Je hebt dan voor hetzelfde werk wel minder analisten nodig met als gevolg dat veel kenniswerkers vervangen zullen worden door AI: die is veel sneller, consistent en goedkoper. Verzekeringsmaatschappijen werken hier allemaal al mee. Hetzelfde gaat gelden voor de advocatuur. Mensen komen niet voor een advocaat, maar voor juridisch advies. En als Watson dat veel sneller, consistent en goedkoper levert, waarom zou je dan nog naar een advocaat gaan. In veel bedrijfstakken zul je zien dat mensen gaan anticiperen op deze vorm van dienstverlening. Daarom is niet te voorzien wat er precies zal veranderen, maar dat de impact groot zal zijn in bedrijfstakken met veel kenniswerkers, staat wel vast. En wat gaan al die kenniswerkers dan doen...?

We zijn in een periode gekomen waarin AI onze samenleving in een rap tempo gaat veranderen. Die ontwikkeling is ingezet en die stoppen we niet meer. Het is nu zaak dat wij bedenken welke rol we aan AI in onze samenleving willen geven. Als we daar nu niet over nadenken, worden we voorbijgelopen door de ontwikkelingen. We moeten nu antwoorden formuleren op de (ethische) vragen die de technologie ons stelt.