

Wat denkt een algoritme van mij wanneer het mij beoordeelt?

Een kritische beschouwing van de inzet van kunstmatige
intelligentie bij beslissingsprocessen

Juni 2020
Symbolische leven I
Jurriaan Duyne - 2697055

Lass uns Mensch sein -- Ludwig Wittgenstein

Inleiding

Dankzij de doorlopende digitalisering van de samenleving komen mensen steeds meer in aanraking met technologie. Laptops, smartphones en in het bijzonder internet is letterlijk niet meer te vermijden zonder jezelf buiten de maatschappij te plaatsen. In het bijzonder is er een interessante plaats gekomen voor kunstmatige intelligentie (k.i.), de nabootsing van menselijke intelligentie. In dit essay wil ik onderzoeken hoe we de technologische inmenging in processen die tot voorheen het gebied van menselijk handelen waren kunnen duiden. Kunstmatige intelligentie bestaat in vele variëteiten en wordt ingezet in diverse toepassingsgebieden: van het vinden van goede zoekresultaten en geven van persoonlijke aanbevelingen tot aan het detecteren van sluipende lekken in pijpleidingen en het ondersteunen of vervangen van beslissingsprocessen. Het laatste punt is het onderwerp dat ik hier verder wil verdiepen.

Om de inbreng van k.i. in de -tot voor kort- exclusief menselijke beslissing te duiden kijk ik naar het 'moderne mensbeeld', dat ik afzet tegen een andere antropologie: die van het ervaren. Voor het moderne mensbeeld gebruik ik Fukuyama's *The End of History and the Last Man* (1993) als archetypische formulering. Specifiek beschouw ik wat *vrijheid* en *vrije markt* betekenen voor Fukuyama en hoe ze de hedendaags dominante blik op de inzet van technologie en het discours daaromtrent kunnen verklaren. Sheikh (2017) volgend stel ik dat onder het moderne denken een een subject-object denken verscholen gaat dat bepaalde gevolgen heeft

voor de duiding van de casus k.i. in beslissingsprocessen. Tegenover het moderne mensbeeld plaats ik, wederom in de lijn van Sheikh, een ervaringsantropologie, met speciale aandacht voor *tacit knowledge* van Polanyi en *Sphären* van Peter Sloterdijk. Vanuit deze twee sterk verschillende perspectieven wordt het vraagstuk van een beslissingsproces geduid.

De methode voor de duiding is het het aandachtig bestuderen hoe een beslissing zich verhoudt tot verantwoordelijkheid in de context van organisaties, hoe de laatste wordt gebruikt en wat voor een symbolische en letterlijke betekenis deze draagt. Het doel van dit essay is om te laten zien hoe het hedendaagse taalgebruik omtrent beslissingsprocessen waarin k.i. wordt gebruikt aansluit of zelfs voortkomt uit het subject-object denken dat het moderne mensbeeld kenmerkt, en om een suggestie te geven voor een alternatief taalgebruik dat past bij een alternatief, naar mijn mening menselijker, mensbeeld.

Twee paradigma's om naar de wereld te kijken

In *The End of History and the Last Man* beargumenteerd Francis Fukuyama dat de strijd tussen verschillende ideologieën is gestreden en dat deze heeft geresulteerd in een overwinning door het liberaal democratisch gedachtegoed. Alle alternatieven, namelijk koninkrijken, feodale stelsel, maar met name de 'recente' bedreigingen van het fascisme en communisme zijn overwonnen. Ondanks dat Fukuyama veel termen in meer of mindere mate gelijkstelt is hij expliciet over de grondbeginselen van *vrijheid*. Voor vrijheid volgt hij Lord Bryce met zijn drie fundamentele rechten: "burgerrechten 'de vrijstelling van van controle over de burger met betrekking tot zijn persoon en zijn bezit', religieuze rechten en politieke rechten" (pp 42-43). Vervolgens ziet hij in het kapitalisme deze vrijheid terug als economisch liberalisme, dat zijn uiting vindt in een vrije markt (p44), waarin private partijen recht hebben op eigen bezit en deze naar eigen inzicht kunnen inzetten.

Bij het gebruik van de term vrijheid valt op dat er telkens wordt geredeneerd vanuit het perspectief van de staatsvorm. Voor de staat is een individu vrij als hij met rust wordt gelaten en zijn eigen bezit mag inzetten voor zijn eigen economische belangen. Verder heeft hij door politieke rechten het recht tot inzetten voor politieke

besluitvorming. Voor de staat is gelijkheid belegd in het feit dat elke burger gelijk is voor de wet, waarin in de liberale democratie aan bovengenoemde rechten is voldaan.

Desalniettemin raakt Fukuyama wel aan het perspectief van het individu in zijn behandeling van de strijd om erkenning van een persoon als meerdere (*megalothymia*) of als gelijke (*isothymia*). In de liberale democratie ziet Fukuyama de manier om de negatieve aspecten van megalothymia te beteugelen door de scheiding van machten en interne competitie en economisering (p190). De economisering slaat zowel op de leidraad voor afwegingen op het niveau natiestaat als op de “college graduate who performs an internal cost-benefit analysis of the career options open to him or her” (p190). Hier gaat het om het liberale aspect van de markt, hoe individuen zelf kunnen beslissen hoe ze economisch deelnemen aan de markt. Toch ziet Fukuyama ook een rol voor de andere erkenningsdrang, die van de van de erkenning als gelijke (*isothymia*). Hierbij noemt hij als de “niet-materiële factoren” zoals waardigheid, respect en zelfrespect. Ook dit zijn aspecten die in een liberale democratie het best tot uiting komen zonder negatieve bijzaken. Helaas is het bij Fukuyama niet helder hoe deze niet-materiële factoren meer diepte krijgen afhankelijk van specifieke culturele contexten en hoe vanuit dit perspectief ook gegarandeerd zou worden dat alle verschillende contexten convergeren naar liberale democratie.

Ik zou hier graag de lijn van Sheikh (2017) volgen om aan te geven dat Fukuyama's droom van een vrije markt en een democratie met vrije subjecten terug te leiden valt naar de droom van Descartes die de mens als “heerser en bezitter” van de natuur wil zien. Vanaf Descartes vormt er in het westerse wereldbeeld een sterk subject-object oppositie denken. De wereld is in de moderne tijd *geneutraliseerd* (p174ff): denk aan landkaarten die gebaseerd zijn op coördinaten die geen intrinsieke waarde hebben, tegenover premoderne kaarten die een godheid als middelpunt van de kaart hadden. Ook zijn andere eenheden geneutraliseerd; ze hebben niet meer het karakter van activiteiten maar van objectief/neutraal meetbare eenheden. Denk aan de invoering van de mechanische klok (tijd in seconden) in plaats van de tijd dat het een arbeider kost om iets te voltooien of afstand (in meters) in plaats van een indicatie van twee weken lopen. We zien dit helder in het vrijemarktdenken: het is niets anders dan het inzetten van neutrale ‘resources’ (p175). De moderne tijd begint

dan ook met de beheersing/overwinning van de natuur; John Locke geeft al aan hoe de natuur bronnen zijn die we kunnen inzetten om waarde toe te voegen met arbeid. Hierdoor herschikken we steeds de wereld naar ons eigen inzicht (p195). Zelfs hoe we onszelf als individu benaderen verandert onder neutralisering. We kijken naar onze talenten als capaciteiten om in te zetten en zien onze opleidingen als economische goederen. We beginnen onszelf steeds meer te duiden door expert-taal die leidt tot een soort afstand, een gevoel van onthechting (detachment) of ontkoppeling.

Als tegenhanger van het subject-object paradigma stelt Sheikh een “niet modern existentiële conditie” die hij aanduidt met *inkapseling* (embeddedness) of de wereld als een “*dwelling*” (verblijfplaats, in de trant waar je aandacht verblijft)(p211). Deze stap komt voort uit het gedachtegoed van denkers na Nietschze (bij wie Sheikh de culminatie van de *materialist man* herkent, zie p208), zoals bijvoorbeeld Heidegger die aangeeft dat er in het subject-object denken geen ruimte is voor de zoektocht naar de betekenis van de modus die het subject-objecten denken is (het subject-object denken blijft vastzitten in wat Heidegger het ontische noemt, het kan nooit vragen beantwoorden over de zin van het zijn zonder meer, welke een ontologische analyse benodigd). Hierbij gebruikt Sheikh de term *afstemming* (attunement) om aan te geven wat er gebeurt: wanneer we engageren in een activiteit stemmen we af met wat voor ons op dat moment relevant is in de omgeving.

Om de inkapseling te duiden maakt Sheikh mede gebruik van Polanyi's *tacit knowledge*. Polanyi (1966) maakt duidelijk dat bij veel activiteiten we duidelijk iets aan het doen zijn, maar niet altijd kunnen uitleggen hoe we dat doen. Onze aandacht is gevestigd op iets relevants voor de activiteit (hiervoor leent hij de term *distal* uit de taalkunde) en de bezigheid wordt uitgevoerd door middel van een praktische handelingen (*proximal*) die een resultaat mogelijk maken. Een mooi voorbeeld is die van het voelen met een blindengeleidestok. Als we proberen te voelen met de stok zullen we onze aandacht in het puntje van de stok verplaatsen en proberen aan te voelen wat voor een soort ondergrond we betreden. Maar als ons wordt gevraagd hoe we dat doen kunnen we niet precies antwoorden. Een precieze fysiologische beschrijving zou interessant zijn, maar is niet equivalent daar het niets zegt over hoe het voor ons voelt ofwel hoe we leren datgene te voelen. Ditzelfde zien we terug in het leren van een sport als tennissen of iets simpels als fietsen. We kunnen de

aandacht wel naar het proximale verplaatsen (de stok in de hand, het racket en de techniek, en een trapper, respectievelijk), maar dit is niet in de aandacht tijdens het uitvoeren van de activiteit en niet altijd relevant tijdens het leren ervan. We zouden dit kunnen doortrekken naar de voor onszelf moeilijk te doorgronden zaken als mensenkennis (wat zien we aan mensen dat ze gevaarlijk of vertrouwd doet lijken) of algemener communicatie.

Een ander aspect waar Sheikh op leunt voor de embeddedness is dat van *Sphären* van Peter Sloterdijk. In navolging van Heideggers *daseinsanalytik*, waarin we in de wereld geworpen zijn, stelt sloterdijk dat we eigenlijk ook altijd in bepaalde situatie geworpen zijn met een sfeer. Er heerst een “bezielde ruimte” tussen de gezichten van geliefden of tussen moeder en kind (Sheikh 2017 p226). We kunnen dit duiden als ervaringen van ruimte die niet neutraal of objectief zijn. We voelen steeds een middelpunt (in een familie, of een land of voetbalteam) waartoe we een verhouding hebben (of het zelf zijn). Er is een grens aan de sfeer waarbuiten het onbekend wordt of we ons in een andere sfeer verplaatsen om ons opnieuw tot de situatie te verhouden.

In het vervolg van het essay zal ik eerst het bereik van de casus van de inzet bij het beslissingsproces afbakenen en verhelderen. Vervolgens wordt met een lopend voorbeeld de duiding van verantwoordelijk binnen een organisatie zo ver mogelijke geanalyseerd, waarbij we de interpretatie van het moderne mensbeeld afzetten tegen die van de ervaringsantropologie, vooral door interpretaties van sfeer in combinatie met tacit knowledge gebruiken om te onderzoeken wat er verandert door de automatisering van beslissingsprocessen.

Kunstmatige intelligentie die mag beslissen

In de alomtegenwoordigheid van telecommunicatieve technologieën is het belangrijk om duidelijk onderscheid aan te brengen in welke positioneringen we tegenkomen ten opzichte van burgers of consumenten. Vaak wordt deze vanuit de technologie ingestoken, zo wordt bijvoorbeeld uitgelegd wat supervised en unsupervised machine learning is, of welke algoritmen gebruikt worden in big data. Ik zou hier liever richten op de functionele aspecten op de grens van de organisatie

en het individu. Ik geloof dat hier ook de grens ligt (zou moeten liggen) tussen de persoonlijke sfeer en de sfeer die bij de buitenwereld -hier specifiek de organisatie in kwestie- hoort.

We kunnen allereerst de inzet van k.i. herkennen die het doel dient om zo goed mogelijk de aandacht te trekken van een individu en deze vast te houden. Denk hierbij aan aanbevelingen (zoekresultaten, advertentie) die gebaseerd zijn op persoonlijk inputs/ online gedrag. Dit kunnen we positief duiden als in het feit dat een goed resultaat een fijne klantervaring oplevert en een klant doet terugkeren, ofwel negatief als het de klant verder manipuleert dan hij of zij goeddunkt. Ik denk hier bijvoorbeeld aan de goede zoekresultaten aan van google, maar ook aan de verslavings tactieken van facebook om alsmaar meer advertenties te laten zien. Ook kunnen we denken aan amazon die suggesties doet voor producten die je mogelijk ook interessant vindt, maar ook aan youtube dat automatisch keuzes laat doorlopen om klanten zo lang mogelijk aan het scherm gekluisterd te houden. Uiteindelijk gaat deze dienstverlening om zoveel mogelijk winst behalen en niet om mensen een zo fijn mogelijke ervaring te geven die ook nog bijdraagt aan een gebalanceerde levensstijl. De acceptatie van deze inzet van k.i. berust mijns inziens op de vrije markt van het moderne denken (we zijn vrije mensen die voor onszelf rationele keuzes maken gebaseerd op feiten - de staat bemoeit hier zich niet mee) en niet op een ervaringsantropologie (is het wel wenselijk/aanvaardbaar dat ik uit mijn attunement wordt verleid?)

Aan de andere kant zien we dat k.i. wordt ingezet om bepaalde beslissingen te maken, ofwel in de ondersteuning van een medewerker ofwel zelfstandig. Het gaat om wat ik hier aanduid met beslissingsprocessen: processen in bedrijven die voorheen de verantwoordelijkheid waren van een medewerker. We kunnen hier denken aan het verlenen van een krediet of een verzekering, of in de publieke sector het toewijzen van een uitkering of zorghulp. Ook vinden we het overal terug bij sollicitatie- en selectieprocedures. De reductie van de menselijke factor bestaat al sinds simpele automatisering, waar met business rules wordt aangegeven waarop een potentiële klant moet worden toegelaten of afgewezen. Hier speelt k.i. juist een steeds bredere rol omdat het nu bijvoorbeeld profielen van kandidaten kan matchen

en zodoende risico minimaliseert, of nu de uiteindelijke beslissing van een mens komt of geheel geautomatiseerd.

In mijn huidige beschrijving van het beslissingsproces zien we de termen van het moderne denken: ik heb het over human resources, de handelingen van de mens worden geautomatiseerd. Blijkbaar kan in deze duiding een bedrijfsproces gereduceerd worden tot een reeks van (min of meer abstracte) handelingen die ook door een computer kunnen worden uitgevoerd. Skidalsky (2015) legt uit dat dit wenselijk is binnen het vrijemarktdenken omdat de computer kapitaal is in plaats van arbeid, en in een vrije markteconomie de voorkeur uitgaat naar waarde creëren door kapitaal (i.p.v arbeid) en dus naar digitaliseren/automatiseren. Dit wordt ook wel de wet van Baumol genoemd. Uiteraard is het vreemd dat tegenwoordig steeds meer dezelfde tactieken worden ingezet bij de overheid, terwijl deze geen economisch belang najaagt. Toch blijkt uit het continu verbeteren/automatiseren en besparen bij de overheid dat ze op dezelfde wijze zijn gaan denken over organisaties: blijkbaar kun je zowel de burger (klant) als de medewerker reduceren tot objectieve zaken waarover je abstracte regels kunt maken die toch nog rijk genoeg zijn om mensen in hun waarde te laten.

Skidalsky duidt automatisering dit als het verdwijnen van *praxis* ten behoeve van meer *phoenesis*. Deze termen komen al van Aristoteles en waar phoenesis activiteit ten behoeve van een duidelijk extern doel is (zoals het schoonmaken van een huis), is *praxis* een activiteit die het doel op zich heeft (een spel spelen is een rechtvaardiging in zichzelf). Toch zou ik willen beargumenteren dat deze duiding nog niet kan geheel verklaren dat klanten bereid zijn om te gaan betalen voor dienstverlening van mensen. Door te stellen dat dienstverleners een doel op zichzelf is nog niet genoeg geduid wat er dan precies leidt tot het verlangen, welke aspecten van menselijke interactie missen en waarom juist die belangrijk zijn. Ik denk dat het belangrijk is om heel precies te bestuderen wat er gebeurt bij een beslissingsproces en hieruit af te leiden wat er mogelijk mist.

Een fenomenologie van verantwoordelijk binnen een organisatie

Om een alternatieve duiding te kunnen geven aan het inzetten van k.i. in een beslissingsproces zullen we eerst moeten verhelderen wat er bij menselijk handelen in hetzelfde proces allemaal gebeurt. We denken vaak binnen organisaties aan organogrammen en diagrammen met zwarte omlijnningen met een paar termen erin de voorstellen hoe de organisatie in elkaar zit. Hier wordt ook bijvoorbeeld gesteld dat er een hoofd van een afdeling (lees: zwart blokje) eindverantwoordelijke is van (het resultaat) van de afdeling. Maar als we kijken naar hoe mensen zich gedragen in een organisatie zien we een ander beeld: veel werkprocessen zijn niet uitgewerkt met vierkantjes en pijltjes en zelfs als ze dat zijn bestaan er nog veel handelingen die historisch zijn ontstaan en simpelweg worden overgedragen op nieuwe werknemers. Daar komt ook nog bij kijken dat er een werkcultuur is binnen een organisatie waarin mensen zich een houding zullen moeten geven. Er is dus vaak een manier van werken die lastig een op een te duiden is met de formele abstracte beschrijving.

In plaats van een procesdiagram te tekenen wil ik hier een beschrijving geven van de ervaring van een werknemer met het zwaartepunt op diens verantwoordelijkheden. Laten we het voorbeeld van een bank nemen. Laten we stellen dat het gaat om een bank waar maar een persoon werkt. Deze is als directeur eindverantwoordelijke over alles wat gebeurt. Deze bank heeft bijvoorbeeld een (start)kapitaal en gaat hier delen van uitlenen tegen een bepaald rentepercentage. Wanneer een gegadigde voor een lening zich aandient zal de directeur een beslissing maken of deze persoon geschikt is voor het uitlenen van krediet of niet. De reden dat hij dit mag uitmaken is omdat hij de verantwoordelijkheid draagt als het misgaat: als de persoon een dief blijkt te zijn zal zijn bank wellicht instorten. Je kunt stellen dat als de twee personen deelnemen aan dezelfde maatschappij hoort hier uiteraard wel bij dat dit binnen de kaders van de wetgeving van de maatschappij gebeurt: de bankier mag niet discrimineren en de kredietnemer moet zijn verplichtingen nakomen. Hierdoor lijkt het een formaliteit die in de bureaucratie de verantwoordelijkheid vervaagt, maar ik zou beargumenteren dat het mensen nog steeds handelen vanuit dit verantwoordelijkheidsbesef.

Voor een deel zien we dit in het woordbeeld van de nederlandse taal. Iemand *draagt* verantwoordelijkheid voor een beslissing of *neemt* zijn/haar verantwoordelijkheid als het misgaat. Het is iets dat we op ons nemen, soms als last maar het kan ook als eer zijn. Als iemand de verantwoordelijkheid voor iets *krijgt* dan impliceert dit ook capaciteit. We spreken erover als over een soort onzichtbaar object dat kan wisselen van een grote last tot een signatuur van erkenning. Tegelijkertijd kunnen we het voelen, wanneer we ons verantwoordelijk *voelen* voor een leed dat iemand is aangedaan. Het houdt altijd verband met dat *onze acties gevolgen hebben voor anderen*. Terwijl we het onzichtbare object van de verantwoordelijk in onze handen hebben of op onze schouders dragen zullen anderen de gevolgen van onze acties merken en mogen zij kunnen ons ter verantwoording *roepen*: we moeten dan verantwoording *afleggen* voor een actie.

Het voorbeeld van de bank was nog te simpel, laten we het uitbreiden. Stel dat onze directeur het te druk krijgt en personeel in dienst neemt. Hij laat op dit moment iemand toe binnen het domein van zijn verantwoording. Immers als de werknemer een vergissing maakt kan het bedrijf omvallen waarvoor hij zelf verantwoordelijk is. Toch draagt hij een stuk van zijn verantwoordelijkheid over op de werknemer die het recht krijgt om autonoom beslissingen te maken. Op het moment dat een werknemer dus iets mag beslissen zal hij of zij spreken uit naam van de directeur en kan daarmee verantwoording moeten afleggen.

Maar wat gebeurt er nu precies op het moment van beslissen? In de pre-digitalisering tijd zou de kredietnemer langskomen op een fysieke locatie. Deze zou inschatten of de locatie overeenkomt met de verwachting van een kredietwaardig instituut als een bank en kijken of het personeel zich juist gedraagt en gekleed is naar verwachting. Met andere woorden deze persoon betreed een sfeer waarin hij of zij zich thuis kan voelen of juist geen deel van uit zou willen maken. Als hij of zij beslist toe te treden tot een gesprek met de kredietverlener signaleert hij/zij dat er een ondergrens van de verwachting is bereikt. Op zijn beurt zal de kredietverlener de potentiële lener inspecteren op gedrag en omgang. In de sfeer van de bank vervullen ze bepaalde rollen en zien van elkaar hoe goed ze deze rollen invullen. Uiteraard zal de potentiële lener bewijsstukken moeten leveren, maar het is moeilijk voor te stellen dat in een gesprek niet een stuk persoonlijke beoordeling plaatsvindt.

Maar wat houdt deze beoordeling nu precies in? Hierboven heb ik een schamele poging gedaan door te zeggen dat personen toetreden in een sfeer en bekijken of dit aansluit bij hun verwachtingen. Dit is geredeneerd vanuit de ervaring met zegt niet precies wat er gebeurt. Ik zou dan ook argumenteren dat dit behoort tot wat Polanyi *tacit knowledge* rekent: Je maakt een beoordeling maar je weet niet precies hoe je dat doet. Er is een *distal* object, de overweging voor een beslissing, met *proximal* kenmerken zoals ik hierboven gaf. Voor een deel is dit, hoe ziet iemand eruit, hoe praat hij/zij, gedraagt hij/zij zich normaal, etc. Deze proximal objecten zijn niet altijd duidelijk voor degene die de beslissing maakt, deze zal misschien zeggen “ja deze persoon komt eerlijk over” of “dit lijkt mij een geschikte kandidaat” zonder te kunnen zeggen hoe ze aan dit gevoel komen.

De gevoelskant is uiteindelijk niet het hele verhaal. Zeker bij het voorbeeld van de kredietverschaffing wordt er normaliter simpelweg een berekening gemaakt met het inkomen of het vermogen van de persoon die een krediet wil. Deze berekening leunt dus zwaar op wat we objectief meetbare ‘eigenschappen’ van een persoon kunt noemen. Bij verzekeringen zullen hier nog vele andere bij kunnen komen zoals het medisch verleden van een persoon of zijn of haar achtergrond, opleidingsniveau etc. Wat we hier in mijn optiek zien is dat de verantwoordelijkheid van de medewerker wordt gemarginaliseerd, ofwel het risico op een foute beslissing wordt geminimaliseerd, door steeds meer nadruk te leggen op de fysieke/objectieve feiten. De medewerker wordt dus meer een administratief medewerker die geen risico draagt. Vanuit de subject-object redenering is dit zeer duidelijk te volgen, immers volgens deze logica is er inherent niet iets fout aan het reduceren van personen tot hun objectief meetbare verschijning. Je zou zelfs kunnen stellen dat het beter is, personen kunnen niet meer discrimineren op bijvoorbeeld huidskleur als dit niet wordt opgenomen als meetbaar feit.

Als we echter nog dieper inspecteren wat er gebeurt bij de menselijke interactie zien we iets gebeuren dat mogelijk het allerbelangrijkste is bij een beslissing of iemand waardig is toe te treden tot een sfeer. Belangrijker dan inkomsten, verleden of andere meetbare feiten is het *aanvoelen* van (afstemmen op) iemands *intentie*. Dit is iets exclusief menselijks: het projecteren van het eigen vermogen om te dromen, om morgen beter te willen zijn dan vandaag, om een

gerichtheid van empathie uit te dragen. Ik bedoel hier te zeggen dat de beslisser probeert de intentie van degene onder de beslissing te vatten. Hij/zij doet nu alsof ze een goed persoon is, maar denk ik dat hij/zij ook tot het einde van onze verantwoordelijkheidsrelatie dezelfde levenshouding zou hebben? Het is een inschatting of een persoon zich in de toekomst wil houden aan de wetten van (zich wil bewegen in de sfeer van) de bank en de maatschappij.

Hier tegenover kunnen we stellen dat een beslissingsalgoritme nooit verder zal kunnen kijken dan wat er objectief te meten valt. Ook kan het nooit verder kijken dan zijn eigen 'training'. Een algoritme kan een nieuw voorbeeld nooit duiden buiten het kader van eerdere voorbeelden dat het zag, het heeft immers een perfect geheugen en geen creativiteit, althans niet in de zin dat het verder dan de objectieve wereld kan kijken. De gerichtheid van een mens is de computer vreemd. Eerlijkheid is hoogstens een niet-lineaire combinatie van gemeten waarden. En deze waarden komen altijd uit iemands verleden, ze representeren wat iemand tot dusver heeft gedaan, zonder dat daar een meting bij zit hoe deze persoon zich daarbij voelt, of wat hij/zij ervan heeft geleerd en hoe in de toekomst de waarde zelf anders zou willen zien. Een mens daarentegen is juist heel gevoelig voor deze instelling. Het zou bijvoorbeeld een zeer cynische hypotheekverlener vereisen om in de aanwezigheid van twee pasgetrouwde, verliefde mensen te stellen dat ongeveer de helft van de huwelijken mislukt en ze dus beter geen hypotheek kunnen nemen.

Het geeft ook aan dat een mens in principe juist in staat is om minder te discrimineren. Waar een algoritme zich alleen maar op fysieke toevalligheden berust kan een mens hieraan voorbij kijken. Stel dat etniciteit meeweegt in het afgeven van een verzekering. Als ik nu een etniciteit heb van een groep die gemiddeld genomen veel gebruik maakt van een bepaalde verzekering dan zou een verzekeringsmaatschappij mij het liefste weigeren op basis van deze etniciteit. Een algoritme met deze 'kennis' zou telkens negatief over mij adviseren. Menselijke medewerkers kunnen zelfs mét deze kennis mij aannemen omdat ze kijken naar en mij beoordelen op mijn intentie.

Conclusie

In het voorgaande hebben we gezien hoe de inzet van k.i. kan worden geduid vanuit het moderne subject-object denken tegenover een ervaringsantropologie. We hebben gezien dat in subject-object termen niet veel tegenwerpingen gedaan zullen worden tegen de inzet van k.i. anders dan dat er niet gediscrimineerd mag worden door een bepaalde vooroordeel door trainingsdata in een algoritme te laten sluipen. Vanuit dit perspectief is de computer neutraal en voor een organisatie een slimme vervanging voor arbeid.

Vanuit het ervaringsperspectief zagen we een andere duiding: we kunnen in de maatschappij discussie voeren over welke objectief meetbare eigenschappen wel of niet meegenomen mogen worden in een k.i. algoritme om het meer of minder gelijk te maken, maar we gaan dan eigenlijk al voorbij aan het feit dat we al gereduceerd zijn tot een meetbaar object zonder menselijkheid in de vorm van intentie. Door door te zoeken achter de verantwoordelijkheid in een beslissingsproces kwamen we erop uit dat een belangrijk onderdeel het vaststellen van intentie is, welke we hebben geduid met tacit knowledge en is verbonden aan sferen. Door deze achterwege te laten ontstaat in mijn optiek een gevoel van vervreemding.

Overigens betekent dit niet dat het zojuist genoemde debat zinloos is. Als k.i. ingezet wordt bij dergelijke beslissingen is het heel belangrijk dat het zo neutraal mogelijk is en hier goed over na wordt gedacht. Overigens is dit ook geenszins een pleidooi tegen k.i. in het algemeen; van het beter reguleren van stroomtoevoer tot aan het uitstippelen van veegroutes op basis van beeldherkenning van vuil op straat, k.i. heeft en zal nog heel erg mooie bijdragen leveren. Ik heb hier echter geprobeerd te expliciteren hoe het beoordelen van mensen op puur objectieve eigenschappen tot een vervreemdend gevoel leidt en dat het debat rijker zou worden als we het over de intentie van mensen zouden hebben: pas dan vatten we echt wat we aan het vervangen zijn met de inzet van k.i. bij een beslissing.

Referenties

- [1] Fukuyama, F. 1993. *The End of History and the Last Man*. Penguin.
- [2] Polanyi, M. 1966. *The Tacit Dimension*. The University of Chicago Press.
- [3] Sheikh, H. 2017. *Embedding Technopolis: Turning Modernity into a Home*. Boom.
- [4] Skidelsky, E. 2015. Hoe automatisering het karakter van werk verandert. In Went, R. et al. *Robot de Baas*. pp 121-134. WRR/Amsterdam University Press