

Robot auto wordt agressief

Artificiële Intelligentie en de Big Tech bedrijven beïnvloeden onze samenleving op een manier waar je vragen bij kan stellen. Is er in de nabije toekomst een app waarmee je je geweten kan meten? Google Ethics?

Robot wordt agressief

“Artificiële intelligentie en de ontwikkeling van sociale robots bieden de mogelijkheid een andere kijk te geven op emoties”, zo schrijft Mariska Kret. Kunnen we blijdschap e.d. programmeren in een robot? Het zou misschien mogelijk zijn maar dan moeten we feitelijk twee vragen oplossen, die elke cel in onze evolutie heeft moeten oplossen: ten eerste de energievoorziening (die staat aan de basis van elke evolutie van een soort, bijv. chloroplasten bij planten). Ten tweede, elke soort kent seksuele selectie (een olifant valt op een olifant en niet op een muis, terwijl de partnerolifant wel ook duidelijk genetisch moet verschillen). Hoe gaan robots dat doen? Feitelijk schets Mariska Kret de robottoekomst nog heel ver weg, ook al stelt ze vast dat die onontkoombaar op ons af komt. Iets verder beschrijft ze met groot enthousiasme dat robots nu al emoties kunnen tonen (simuleren?) die niet voorgeprogrammeerd zijn. Zij beschrijft het videofragment waarin twee robotauto's in een te kleine ruimte rondrijden. “Er ontstond natuurlijke competitie, waarin de ene auto de andere auto aanviel en wegduwde. Het interessante aan dit agressieve gedrag was dat het ontstond en niet voorgeprogrammeerd was: de twee auto's waren identiek toen ze vers van de pers kwamen.” [01] Dit fragment bewijst des te meer dat we het over de ethische aspecten van Artificiële Intelligentie (AI) en robotologie moeten hebben.

Ethiek op dit punt van samengaan van techniek en mens en de wederzijdse beïnvloeding, is niet nieuw. Sinds de technische revolutie heeft de mens een ambivalente, verwarrende en ingewikkelde verhouding met technologie en machines c.q. robots. Onze oordelen over de technische vooruitgang schommelen tussen utopische en dystopische opvattingen. [02] Met enige trots markeer ik mijn persoonlijke positie t.o.v. de nieuwe media inzake de ethische aspecten in het jaar 1998. In dat jaar las ik het boek ‘Media Literacy’ van W. James Potter, waarin hij een stevig pleidooi houdt om in onderwijsinstellingen een nieuw vak te introduceren, namelijk Media Geletterdheid. Basiselementen hierin zijn allereerst dat een ieder moet weten hoe de menselijke geest werkt en hoe je

persoonlijke kennisstructuren in elkaar zitten. Vervolgens is kennis nodig van elke mediaverschijning, zodat je weet hoe iets functioneert. Als laatste komen deze twee factoren bij elkaar in het verkrijgen van inzicht wat de effecten zijn van al deze mediaverschijningen. En zonder enige trots meld ik dat tot aan mijn pensioen in 2020 mijn kruistocht voor zo'n vak in het HBO en middelbaar onderwijs volstrekt zonder resultaat is gebleven.

Overstijgen van persoonlijke ethiek

Meer en meer wordt duidelijk dat nieuwe media en de Big Tech bedrijven invloed hebben op de individuele mens. Twee voorbeelden van individuele effecten van technologie illustreren dit: allereerst de anti-spiek software voor tentamens op de universiteit. Deze software herkent personen met een donkere huidskleur niet of erg moeilijk. Als de student na vele pogingen eindelijk is ingelogd bij het tentamen, gebeurt het soms dat zo'n student plots uitgelogd wordt ‘omdat de software denkt dat de student uit beeld is’ (die kan dus mogelijk met iemand overleggen etc.). Berucht is in 2015 het feit dat Google Photos zwarte mensen aanmerkte als gorilla's. [03] Ten tweede: Big Tech bedrijven willen niet alleen ons gedrag in kaart brengen maar ook binnentreden in onze gedachten en gevoelens. Zij willen ons manipuleren met individuele (politieke) boodschappen, reclame, en zo binnentreden in onze gevoelswereld. Het blijkt dat de individuele persoonlijkheid van facebookgebruikers redelijk goed afgeleid kan worden uit de duimpjes die de gebruiker bij berichten opsteekt. [04] Big Tech hervormt hele sectoren van de samenleving en overstijgt zo het hierboven beschreven individuele niveau. Hele sectorale disrupties veranderen bestaande structuren met een knip van de vingers: televisie imperium (alles online), taxiwereld (Uber), restaurants (online bestellingen), muziekindustrie (Spotify), bankwezen (bitcoins), telecomindustrie (skype) etc. [05] Zeer beangstigend is de technologische ontwikkeling die binnen de Big Tech bedrijven gerealiseerd wordt. Hun computers m.b.t. AI zijn zo krachtig geworden dat universiteiten niet meer mee kunnen komen. Hun budget- en menskracht staat niet in verhouding: de ‘vertaalgroep’ aan de Universiteit van Utrecht telt 15 mensen en is een van de grootste universitaire vertaalgroepen in de wereld. Bij Google telt de vertaalgroep inmiddels honderd leden. En het algoritme van Googles

nieuwe chatbot, LaMDA, is veel te groot voor de computers van een doorsnee universiteit. Er dreigt een tweedeling in de wetenschap, zij die werken op de universiteit en zij die werken bij een van de Tech giganten. [06] Het besef dat technische artefacten niet alleen invloed hebben op ons persoonlijke leven maar ook op de inrichting van de samenleving als zodanig, dringt steeds meer door. Niet vreemd dat het boek van Frits Bolkenstein e.a. dat gaat over de verwevenheid van ethiek en technologie de titel draagt: ‘De politiek der dingen’. Bolkenstein karakteriseert de politiek als ‘toegepaste ethiek’. De Amerikaanse techniekfilosoof Langdon Winner schreef in 1980 een klassiek geworden artikel, getiteld “Do Artifacts have Politics?”. Daarin staat de roemruchte beschrijving van de racistische autoviaducten over de parkwegen op Long Island. De New Yorkse bouwmeester en planoloog Robert Moses ontwierp die viaducten met een extreem lage doorrijhoogte, n.l. 2.7 meter, zodat de arme zwarte bevolking (die geen personenauto konden bezitten) er niet met de stadsbussen onderdoor konden gaan. (zie afbeelding) De stranden bleven exclusief terrein van de rijke witte bevolking die met hun personenwagens het prachtige strand wel konden bereiken. De anekdote wil dat vandaag de dag de rijke witte klasse allemaal hoge kampeerwagens bezitten en nu dus zelf niet meer het strand en dat deel van de stad kunnen bereiken. [07]

Tijd voor ethiek

We starten, misschien wel iets te persoonlijk, met een jeugdboek van mijn vader “Slechte tijden” door Charles Dickens in een prachtige Jugendstil kaft. De hoofdpersoon is leraar Thomas Gradgrind, “een man van het positieve en materiele. Een man van feiten en cijfers. ... Met eene lineaal en een goudschaaltje en de tafel van vermenigvuldiging altijd in zijn zak, gereed om ieder staaltje van het mensdom te meten en te wegen. ... Het was een grondregel der Gradgrindsche filosofie, dat alles betaald moest worden. Niemand mocht ooit om de eene of andere reden iemand iets geven of eenige hulp bewijzen, zonder dat er voor betaald werd. Dankbaarheid moest afgeschafte worden, en alle deugden, die daaruit voortsporen, behoorden niet te bestaan.” [08] Alsof Dickens onze huidige neoliberale samenleving beschrijft avant la lettre. Maar Dickens kritiseerde in deze, voor zijn doen, korte roman de opkomst van de Industriële Revolutie en



de daardoor ontstane enorme verschillen in rijkdom tussen mensen. En hij poogde tevens de toen opkomende filosofische stroming, het utilisme (= nuttigheid is het enige richtsnoer) van Jeremy Bentham (1748-1832) en John Stuart Mill (1806-1873) belachelijk te maken. Bentham kritiseerde elke ethische theorie die een objectief rechtvaardigheids criterium als grondwet voor ons handelen positioneerde. Hij vertrekt vanuit de psychologie en stelt dat elk mens onderworpen is aan ‘twee soevereine meesters’, namelijk pijn en genoegen. Elk mens vermijdt pijn en probeert zoveel mogelijk genoegen te genereren. Om onze samenleving dus goed te laten verlopen is het zaak dat elke beslissing die we nemen tegen deze twee, pijn en genoegen, wordt afgezet. En van elke beslissing kunnen we een berekening en wegging maken zodat de handeling het grootste geluk voor het grootste aantal mensen voortbrengt. “Bij de keuze tussen alternatieven is de hoeveelheid plezier het enige criterium”. Dat kan tot vreemde situaties leiden zoals filosoof Carritt betoogt. Stel een moordenaar wordt opgehangen omdat zo eventuele moorden in de toekomst niet meer door hem/haar gepleegd kunnen worden. Het utilisme keurt in feite deze ophanging goed. Het criterium ‘het grootste geluk’ leidt soms tot niet wenselijke morele daden met soms absurde gevolgen. [09] Hoe onethisch de ethiek van het utilisme ook lijkt, de filosoof Ad Verbrugge memoreert wel dat dit gedachtengoed ‘de wending naar lichamelijke behoeftes in de moderne filosofie’ veroorzaakt heeft. Bentham zorgde tevens voor betere omstandigheden in de Engelse barbaarse gevangenissen en zo meer. Misschien verschilden Bentham en Dickens feitelijk niet zo heel veel. [10]

Toekomst: Google Ethics

Het is niet een vreemde associatie om AI en het utilisme met elkaar te verbinden want AI zou inderdaad een hyper snelle be-

rekening kunnen maken tussen kwantificeerbare geluks- en genotselementen in elke ethische beslissing. AI kan doeltreffend uitrekenen hoeveel mensen positief vooruitgaan bij een besluit en hoeveel er mogelijk onder gaan lijden. De uitkomst is objectief en kwantificeerbaar: het meeste geluk voor een zo groot mogelijke groep mensen. Google Ethics en Dickens leraar Thomas Gradgrind zijn zo gek nog niet.

De filosoof Guido van der Knaap heeft hier nou net in 2022 een boek over geschreven: Van Aristoteles tot algoritme (filosofie van kunstmatige intelligentie). Laten we daar eens mee beginnen op de andere pagina van deze krant. Is Google Ethics verre toekomstmuziek of schieten filosofen dit bizarre idee aan flarden?

Reageren

rikroelfzema@gmail.com
Dit artikel digitaal?
www.rikroelfzema.nl
(klik op ‘Over mijn publicaties’)

Dit artikel is gebaseerd op de volgende boeken:

- [01]= Tussen glimlach en grimas (uitingen van emotie in mens en dier) (2022), Mariska Kret, blz. 189 + 200
- [02]= Leven in Media (2018), Mark Deuze, hfst. 13: Media maken ons zowel tot slaven als dat ze ons bevrijden.
- [03]= Bewijs maar dat je een mens bent (Volkskrant 15 juli 2022) (reportage over antispieksoftware)
- [04]= Big Tech bedreigt de vrijheid om te denken wat je wil (NRC 4 juli 2022) (interview met Susie Alegre)
- [05]= Me the Media (verleden, heden en toekomst van de derde mediarevolutie) (2008), Jaap Bloem, Menno van Doorn, Sander Duyvestein. Hoofdstuk 8: Vijf sectorale disrupties en een thematische
- [06]= Wetenschappers zijn bang voor hun toekomst door Googles superieure computer (Het Financieel Dagblad 2 juli 2022)
- [07]= De politiek der dingen, Frits Bolkenstein e.a. (2009). Enkele hoofdstukken zijn erg interessant: ‘Elektronische dossiers en echte kinderen (waarden en ict-systemen in de jeugdgezondheidszorg)’, of ‘Wie zoekt zal vinden? (de politiek van zoekmachines op internet)’, of ‘Een blik in de baarmoeder (de geopolitiek van de ecoscopie)’ om er een paar te noemen.
- [08]= Slechte tijden, Charles Dickens (z.j. ongeveer eind 1800), blz.2 + 138. Dankjewel vader!
- [09]= Geschiedenis van de ethiek, Alasdair Macintyre (1966)(1974), blz. 234-255
- [10]=Het goede leven & de vrije markt, Ad Verbrugge e.a. (2018), blz. 236 e.v.

Google Ethics weegt uw geweten

Als alles te herleiden is tot data, wat is dan 'rechtvaardigheid' bijvoorbeeld? Is er een theorie die de AI en Big Tech onder ethische controle kan zetten?

Surveillance-kapitalisme

"Harvard-hoogleraar Shoshana Zuboff noemt de werkwijze van grote technologiebedrijven zoals Google surveillance-kapitalisme. Aan de basis van deze vorm van kapitalisme liggen grote hoeveelheden data waarmee het mogelijk is om gebruikersprofielen op te stellen, gedrag te voorspellen en gepersonaliseerde diensten aan te bieden." [01] Het spel Pokémon GO is een perfect voorbeeld. Het lijkt alsof kinderen Pokémons moeten vangen maar dat is een dekmantel want feitelijk kopen bedrijven Pokémons in, om de kinderen naar hun winkel, café of restaurant te lokken. Bovenstaande beschrijft filosoof Guido van der Knaap en we gebruiken zijn boek om eens dieper in de ethiek achter de AI en de Big Tech bedrijven te duiken. Allereerst serveert hij het utilisme (zie andere pagina) af. De kille optelsom van de hoeveelheid genot negeert onze diepste menselijke drijfveren waarom we iets doen. Het negeert dus onze integriteit. "Het wrange is dat utilitarisme draait om handelingen met de juiste gevolgen, maar geen ruimte biedt voor wat ons tot handelen aanzet. Of dat nu liefde voor familie en vrienden is, of een passie voor kunst." [02] Van der Knaap zoekt zijn heil in de plicht-ethiek (deontologie, deon = plicht) van Immanuel Kant. Hier gaat het om wat iemand doet en waarom, los van de gevolgen (die in het utilisme juist alleen belangrijk zijn). Ethisch handelen doe je volgens een plicht die voor iedereen hetzelfde is. Je mag niet stelen. Dat geldt voor jou en iedereen.

Zo voorkom je willekeur want het zou vreemd zijn dat een ander dat plots wel zou mogen. En, stelt Kant, je moet de plicht autonoom uitvoeren. Jij bepaalt dat je je aan die plicht wilt houden. Dat betekent dat je wel inzicht moet hebben in het eigen handelen. En dit is wat AI en de Big Techbedrijven niet met ons delen: autonomie. Achter onze rug bepalen zij voor ons (bijv. welke Netflix serie we gaan kijken, welk nieuws we ons voorschotelen, hoe we via Google Maps ergens komen). "Persoonlijke zoekresultaten of nieuwsberichten geven ons de illusie dat wij een onafhankelijk totaaloverzicht krijgen, terwijl ze op ons eigen zoekverleden zijn toegesneden. Daardoor kooft AI steeds meer beslissingen voor ... Ze pakt ons brein af en onze autonomie verdwijnt. ... Wie de brievenbus opendoet en daarin een al geopende envelop aantreft, is op zijn minst verontrust. ... Bij het aftappen van e-mails of chatberichten is dat veel minder tastbaar" [03] Dan probeert Van der Knaap het nog een derde en laatste keer: is AI binnen een ethisch systeem te plaatsen? Hij haalt Aristoteles erbij met zijn deugdethiek. [04] Deugden zijn karaktertrekken die een mens in zijn leven door vallen en opstaan ontwikkeld heeft. (Dat zou AI ook kunnen!) De vier belangrijkste zijn moed, gematigheid, verstandigheid en rechtvaardigheid. En elke deugd is eigenlijk de gulden middenweg tussen een teveel en te weinig: te weinig moed leidt tot lafheid maar een teveel aan moed leidt tot overmoed. (Deze menselijke afweging in het hier en nu wordt al moeilijker voor AI.) Dan geeft Aristoteles de laatste voorzet die dodelijk is voor AI. De deugd zowel als de ondeugd is door de mens vrijwillig gekozen en hangt ook volledig af van die mens zelf. Het is jouw geweten

(nl. welke plichten je wel of niet gestalte wil geven en hoe). Google Ethics kan deze verantwoordelijkheid tot zelfstandig kiezen dus nooit van ons overnemen.

Techniek werkt aan frictieloosheid

In haar boek 'Frictie' onderzoekt Miriam Rasch de ethiek in tijden van dataïsme. Weinig goeds voorspellend schrijft ze aan het begin van haar zoektocht dat 'dit boek zich moeilijk laat navertellen'. [05] In onze huidige samenleving lijkt de techniek het gehele leven terug te brengen tot data: hoe meer data, hoe beter de voorspellingen en oplossingen van die techniek zullen zijn. Een frictieloos bestaan is het uiteindelijke doel van de techniek: alles opgelost na het verzamelen van oneindige data waarmee AI de finale oplossing voor alles biedt. Moest je vroeger nog bewijzen dat je geen robot was door stoplichten, huisnummers of brandkranen te tellen (Google Maps heeft je er nog nooit voor bedankt terwijl je ze onbewust hielp), zo is nu nog slecht het aanvinken van het hokje 'ik ben geen robot' voldoende. "Dat je daarmee naar waarheid handelt, leidt Google af aan de bewegingen die je maakt met je muis naar het hokje toe, je IP-adres, cookies – de hele context waarbinnen de beweging van het vinken zich afspeelt." [06] Minder frictie ook hier: onbewust heb je zelf het probleem of je een robot bent opgelost. En in dat dataïsme ligt wel een probleem: de mens hoeft zelf geen beslissingen meer te nemen want de data hebben eerder in de gaten wat er met je gaande is dan jij zelf (bijv. analyse van je pupil bewegingen maakt duidelijk wat je echt wilt hebben). Helaas moet veel techniek nog heel wat leren want er ontstaan nogal wat oneffenheden. Zo herkent de zeepdispenser niet een zwarte hand, dus geeft geen zeep (foutje van het algoritme) of het algoritme van Facebook haalt de foto van het naakte Vietnamese meisje dat door napalm verbrand is, offline. Naakt dus delete. Het zijn voorbeelden van algorithmic biasen, vooroordelen van de dataverzameling. Aan de basis van het dataïsme ligt een fundamentele vraag: wat beslissen we nog zelf? De data worden geanalyseerd en er volgt gelijk een uitkomst, er is niet meer het moment van de beslissing. De beslissing is het systeem zelf en kunnen we onmogelijk achterhalen. Voorbeeld: je wil een lening aanvragen. De bank analyseert een oneindige reeks van data over jou en vervolgens word je geconfronteerd met de uitkomst, wel of geen lening of tegen onduidelijke meer kosten want er blijken voor de data risico's aan je te kleven. "De fixatie op getallen brengt echter haar eigen logica met zich mee. Ze schept haar eigen

wereld". Rasch geeft als voorbeeld hoe de klant het personeel die hem geholpen heeft, moet ranken met een getal van 1 tot 10. "Degenen die door cijfers beoordeeld worden, zullen hun energie stoppen in het tevredenstellen van de cijfers, ook als ze daarvoor met de cijfers moeten sjoemelen". [07] Het doel is middel geworden. De persoon die wij worden in een dataverzameling gemaakt door digitale technologieën stemt vaak niet overeen met hoe wij onszelf ervaren. Er is sprake van 'ethische differentie' (onemigheid tussen mijn zelfbeeld en wat er van mij gemaakt wordt in analyses van mijn data).

Verzet tegen de Big Tech

Wat is het antwoord van Rasch m.b.t. de ethische differentie? De Big Tech lijken onverslaanbaar. Rasch positioneert de mens als 'de-automaton'. De mens kan de dataverzamelingen disrupten, verstoren. Maar het hoofdstuk hierover telt maar zeven pagina's. [08] Worden we dan iets wijzer? We wenden ons tot de laatste, sterk poëtische hoofdstukken. Rasch stelt dat de dingen bestaan in de dataverzamelingen maar dat de dingen zelf eigenlijk groter zijn dan de dataverzameling. De dingen zelf lijken te ontsnappen. (In kunst is dit goed te begrijpen: de dingen zijn nooit wat ze lijken te zijn maar ze zijn altijd meer, soms een voertuig van een idee of gevoel of ook weer niets zoals in het schilderij 'Het Zwarte Vierkant' uit 1913 van Kazimir Malevitsj. Het schilderij noemde hij de vormgeving van de absolute nul. Zelf, zo zei hij, bleef het schilderij een eeuwig raadsel voor hem.)

Rasch zoekt hulp bij de filosofische stroming de object-georiënteerde ontologie (ooo). De ooo-filosoof Graham Harman definieert de wereld als een totaal aan ontelbare objecten, onmeetbaar en ieder met autonomie en eigenheid. Ook de mens is simpel object, niet meer of minder dan de appelboom of het doosje lucifers. De objecten zijn niet reduceerbaar tot iets anders, dus ook niet tot data. Ons denken lijdt aan vier denkfouten (valkuilen) waardoor de dingen feitelijk aan ons lijken te ontsnappen. De eerste valkuil is dat we alles willen terugbrengen tot tastbaarheid, met afmetingen en herkenbare eigenschappen. Dit noemt hij de valkuil van het fysikalisme. Hij geeft als voorbeeld De Vereenigde Oost-Indische Compagnie. Die is enerzijds tastbaar en ook abstract. De tweede valkuil is onze voorkeur voor het kleine:

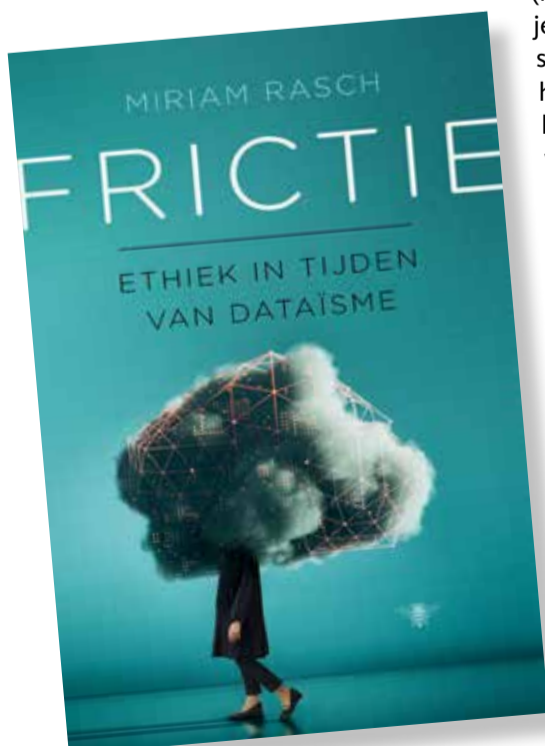


de VOC terugbrengen tot x schepen, y bestaansjaren etc. De derde noemt hij onze voorkeur voor realisme. Sherlock Holmes bestaat, hij is reëel. Dat klopt maar hij is ook tegelijk fictief. De vierde valkuil is de grootste: wat bestaat kan worden uitgedrukt. Maar, zo stelt hij, de dingen ontsnappen aan ons in een beschrijving want dat is slechts een vertaling van het ding en niet het ding zelf. Daarom kunnen we maar beter accepteren dat we over de werkelijkheid slechts in metaforen kunnen spreken. "De metafoor is niet geïnteresseerd in de dimensies of deeltjes van een object, in realisme of definieerbaarheid. Ze gaat niet uit van bestanddelen, noch van effect." [09] Dit is de wraak van de dingen waaraan we als mens enkel gelijk zijn. "Dat geldt ook voor de techniek zelf. Uiteindelijk, als alle dingen gelijk zijn, even onkenbaar maar evengoed allemaal te belichamen, dan zijn data, het internet of algoritmes dat ook. Ook die zijn immers objecten, ondoordringbare grootheden ... die zich nooit echt van hun binnenste laten zien." "Het echec van de mens – dat er ... altijd onzekerheid en duisternis in haar huizen – is net zo goed het echec van data. Ook in de werkelijkheid geproduceerd door getallen en binaire code, een werkelijkheid die zo graag totaal en alomvattend zou zijn, zit iets wat aan zichzelf ontsnapt, ook het data-object is een instantie van het verborgene." [10] Mooie poëtische woorden maar of een mens hier iets mee opschieft? Of de Big Tech zich hierdoor laat afschrikken?

Reageren

rikroelfzema@gmail.com
Dit artikel digitaal?
www.rikroelfzema.nl
(klik op 'Over mijn publicaties')

Dit artikel is gebaseerd op de volgende boeken:
[01]+[02]+[03]+[04]= Van Aristoteles tot algoritme (filosofie van kunstmatige intelligentie), Guido vd Knaap (2022), blz.188; blz.144; blz.149+151; blz.154-157.
[05] t/m [10]= Frictie (ethiek in tijden van dataïsme), Miriam Rasch (2020) (2021), blz.14; blz.49 e.v.; blz.111; blz. 73-79; blz.178, blz. 188+198.



Ethische handleiding voor techniek

Dat robots en artificiële intelligenties in de toekomst onze medemens zullen zijn, dat staat vast. Maar wat zal de interactie tussen mens en algoritme / robot zijn? Een zegen of een hel? Nu nadenken over ethiek is noodzaak.

Surveillance-kapitalisme

De historicus Yuval Noah Harari is een transhumanistisch denker. Transhumanisten gaan ervan uit dat het onvermijdelijk is, dat er een algemene artificiële intelligentie komt die de mens zal overschaduwen. Dat scenario kan apocalyptisch of eschatologisch (= einde der tijden) zijn en dus onze wereld definitief ontworpen. Bij dit doomsce- nario hoort onze angst voor robots die door de sciencefiction schrijver Isaac Asimov 'het Frankensteincomplex' wordt genoemd. Mary Shelley schreef het boek 'Frankenstein', waarin een doodpersoon m.b.v. een wetenschappelijk experiment weer tot leven werd gewekt (en met desastreuse gevolgen).

Harari stelt: "De belangrijkste economische vraag van de eenentwintigste eeuw zou wel eens kunnen worden wat we aan moeten met alle overbodige mensen. Wat zullen bewuste mensen doen zodra we uiterst intelligente, niet-bewuste algoritmen hebben die bijna alles beter kunnen?"[01] Computers maken veel menselijke arbeid overbodig, ze doen het sneller, beter en goedkoper. Harari gaat nog verder door een onderzoek van Facebook te citeren waarin bewezen wordt dat Facebook betere voorspellingen over het karakter en de persoonlijkheid van een persoon kan doen, dan vrienden, ouders of partner. Het Facebookalgoritme kon zelfs op sommige punten (bijv. drugsgebruik, sociale relaties) een betere evaluatie geven dan de persoon over zichzelf kon doen. Dat Facebookalgoritme is niet zonder gevaar want dat weet precies wie en waar en waarmee de zwevende kiezer in bijv. de Amerikaanse verkiezing bestookt moet worden. En als dat Facebookalgoritme ook nog een persoonlijke politieke voorkeur zou bezitten, bijvoorbeeld in de persoon van de directeur van Facebook, dan is de noodzaak van een ethiek niet ver weg. "De nieuwe technologieën van de 21ste eeuw zouden dus de humanistische revolutie teniet kunnen doen, mensen hun autoriteit kunnen ontnemen en niet-menselijke algoritmen aan de macht kunnen brengen."[02] Vermoedelijk zal slechts een beperkte groep mensen toegang hebben tot de techniek die ons menszijn een upgrade kan geven m.b.v. (bio)technische innovatie door AI (chips in ons



brein, medische controle en versterking van ons lichaam). Maar wat zal de plaats van de (elite)mens zijn in de oceaan aan data? "Het individu wordt steeds meer een klein chipje in een gigantisch systeem dat niemand echt begrijpt. Ik absorbeer dagelijks talloze databits via mailtjes, telefoontjes en artikelen. Ik weet niet echt wat voor plaats ik inneem in het grotere geheel of wat voor verbanden mijn stukjes data aangaan met de data die miljarden andere mensen en computers produceren."[03]

Maar ook de ontwikkelaars van dit alomvattende Facebook dataisme weten niet echt waar ze mee bezig zijn. Elk lid uit zo'n gigantisch ontwikkelteam "kent maar één stukje van de puzzel en niemand begrijpt het hele algoritme helemaal. Bovendien komen er met de opkomst van 'machine learning' en kunstmatige neurale netwerken (= computer / robot leert zelfstandig door AI) steeds meer algoritmen die zelfstandig evolueren, zichzelf verbeteren en leren van hun fouten. ... naarmate het algoritme groeit, gaat het zijn eigen weg en betreedt het paden waar geen mens ooit is geweest – en waar geen mens ooit zal kunnen komen."[04]

Noodzaak van ethiek

Of Harari een toekomstige utopie of een apocalyptische en eschatologische desastreuse eindtijd schetst, dat is niet helemaal duidelijk. Maar dat we over die verre toekomst ook ethisch zouden moeten reflecteren, dat lijkt mij duidelijk. Hoe goed bedoeld alle 'ethische commissies' in grote steden ook zijn, ze bespiegelen eigenlijk alleen dat wat reeds bestaat (bijv. is

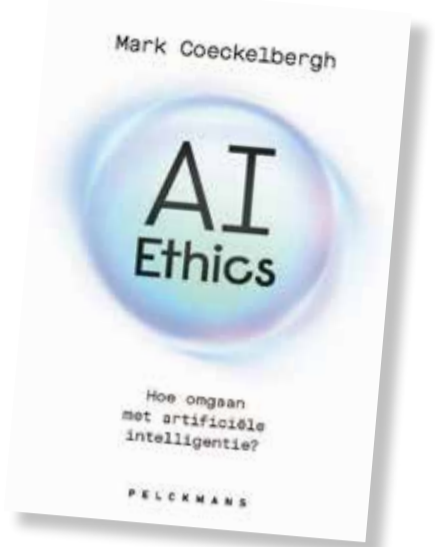
de body Cam van Boa's wel gewenst? De onschuldig gefilmde passant!) en in de nabije toekomst gerealiseerd kan worden (mogen de corona-druktemeters in het stadscentrum nu voor iets anders gebruikt worden? Ze hangen er immers). [05] Gelukkig schetst techniekfilosoof Mark Coeckelbergh een echt vergezicht m.b.t. Artificiële Intelligentie. Hebben robots en superintelligente AI's morele status, of zijn ze slechts dode dingen? (Zijn ze dus moral agents, in staat tot morele, bewuste en autonome keuzes?). "Zelf rijdende auto's....Stel dat er voetgangers de weg op rennen en de auto die alleen kan ontwijken door tegen een muur te rijden en de chauffeur op te offeren. Wat moet de auto dan doen?" [06] Kunnen mensen wel in zo'n split of a second goede morele beslissingen nemen, of kunnen AI's dat beter wegens hun razendsnelle berekeningen m.b.t. de ontstane fysieke situatie? Wat te denken over kwesties als 'vergroot AI de afstand tussen arm en rijk', 'hoeveel banen gaan door AI verloren', 'bevoordeelt of discrimineert AI', 'wat is de ecologische voetafdruk van AI', 'hoe staat het met onze privacy en tot hoever worden we gecontroleerd', 'in hoeverre zijn sociale robots een illusie m.b.t. vriendschap', 'zijn zwakbegaafden en kinderen niet te gemakkelijk te manipuleren door AI', 'wat als de software gehackt wordt van robots', 'mag een AI robotsoldaat schieten op alles wat leeft?' Coeckelbergh behandelt de kwesties hoofdstuk voor hoofdstuk. Essentiële vragen zijn hierbij of de mens mee moet kunnen kijken in al dit soort processen. Dat zal onmogelijk zijn want de

mens heeft maar een beperkt denkraam t.o.v. AI. Moeten AI-beslissingen altijd transparant en verklaarbaar zijn? Dat is onmogelijk want aan AI werken duizenden programmeurs met ieder een klein stukje kennis van de puzzel (het algoritme) en het is dus onmogelijk om elke handeling van een AI te voorspellen of te kunnen uitleggen. Daarbovenop: willen de bedrijven hun algoritmes wel prijsgeven want dat is immers hun investering, de intellectuele eigendomsrechten. "We kunnen concluderen dat ethische discussies rond AI ons doen nadenken over hoe een goede, eerlijke samenleving er zou uitzien, wat een mensenleven zin geeft en welke rol technologie in beide vraagstukken speelt (of zou kunnen spelen)."[07]

Mens voorrang boven techniek

Het lijkt makkelijk: de mens heeft de hoogste rechten. Toch toont onze filosofische geschiedenis dat we bijvoorbeeld anders over de natuur en het leven daarin, zijn gaan nadenken. De mens is niet meer vanzelfsprekend de koning op de ethische troon. We hebben een steeds kritischer visie op antropocentrische denkpatronen, de mens staat niet vanzelfsprekend meer in het centrum van het leven. De vraag is dus: hebben robots en superintelligente AI's recht op een eigen bestaan, kortom moeten mensen hun bestaan respecteren (zijn zij moral patients?). [08] Een ontwikkelde superintelligente AI: "Is het dan ethisch verantwoord om die uit te schakelen, om die te 'vermoorden'? Is het oké om een robohond een schop te geven?"[09] Vanuit de deugdedethiek dienen we de AI's met respect te behandelen want doen we dat niet dan glijden we af in ons morele gevoel, volgens Kant dan. Schoppen tegen de robots leidt automatisch naar de glijdende schaal om ook mensen te schoppen. De deugd 'respect voor een ander wezen (wat dat wezen dan ook is)' staat onder een niet te accepteren druk. Het grootste gevaar is dat we al deze ethische vragen niet stellen omdat we denken dat de techniek in de toekomst alles zal oplossen. "Technosolutionisme: de overtuiging dat we voor elk probleem een technologische oplossing kunnen vinden en dat niets onze menselijke en artificiële intelligentie te boven gaat."[10]

Een handleiding, die is er Maar heeft Coeckelbergh een ethische handleiding geschetst? Ik vrees: niet echt. De filosoof Eynikel doet dat veel concreter.



Hij schets een innovatieproces dat altijd waarden gedreven moet zijn. Allereerst is er een analyse nodig wie betrokken is bij een innovatie: direct in contact met de innovatie en haar effecten, of indirect beïnvloed erdoor (= MENS). Als tweede moet er een impactanalyse komen van de gevolgen van de innovatie (worst- & best case scenario; de hard impact en de soft impact).(=MAATSCHAPPIJ). De laatste stap is de waarden- toets (=MORAAL): welke waarden worden verbeterd en welke verslechterd door de innovatie. Denk hierbij aan de uitvinding van de auto want die draagt bij aan de waarde 'vrijheid' maar knaagt aan de waarde 'gezondheid'. Of neem de slimme energiemeter, die is goed voor het milieu maar knaagt aan ons gevoel van privacy. Hulpmiddelen daarbij zijn de technische bijsturingen van de innovatie om negatieve zaken te elimineren en de morele omkadering (= welke mensen, partijen, groepen kunnen ons helpen en hoe krijgen we ieder op een lijn) om gezamenlijk de innovatie te dragen. Misschien is zijn 'mensmaatschappij-moraal'- model zo gek nog niet. Concreet en draagvlak zoekend.[11] Er is nog zoveel te doen. Wachten en vrijblijvend filosoferen is onverstandig. Wie pakt de handschoen op?

Reageren

rikroelfzema@gmail.com
Dit artikel digitaal?
www.rikroelfzema.nl
(klik op 'Over mijn publicaties')

Dit artikel is gebaseerd op de volgende boeken:

[01] t/m [04]= Homo Deus (een kleine geschiedenis van de toekomst) Yuval Noah Harari (2015)(2018), blz. 328; blz.354-355; blz.397; blz.405
[05]= Trouw 6-8-2022 Kunstmatige Intelligentie (er is behoefte aan ethische reflectie), pag. 8-9, bijlage religie&filosofie
[06]+[07]= AI ethics (hoe omgaan met artificiële intelligentie?), Mark Coeckelbergh (2020)(2022), blz.46; blz.117;
[08]= De journalist en informaticus Henny van der Pluijm publiceerde in 2013 zijn boek 'Rechten en plichten voor robots', waarin hij de toekomstige mensenrechten voor AI schetst en probeert in te vullen. Zeer lezenswaardig, hier te weinig ruimte om op in te gaan.
[09]+[10]= AI ethics, blz.48; blz.163
[11]= Robot aan het stuur (over de ethiek van techniek), Jochanan Eynikel (2017), deel 3, hfst.2 blz.113-139