



SOGETI

VINT | VISION • INSPIRATION • NAVIGATION • TRENDS

MACHINE INTELLIGENCE

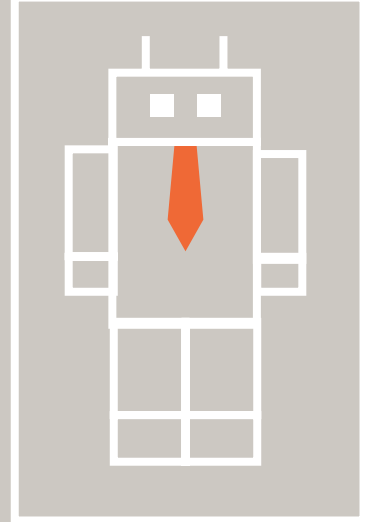
3/4



De FrankensteinFactor *Anatomie van de angst voor AI*

Menno van Doorn, Sander Duivestijn, Thijs Pepping

INHOUD



We danken het Oxford Internet Institute (OII), voor de inzichten en ideeën die kwamen uit een gezamenlijke Machine Intelligence-workshop, met name: Luciano Floridi (Director of Research, Professor of Philosophy and Ethics of Information), Scott Hale (Data Scientist), Vili Lehdonvirta (DPHil Programme Director, Research Fellow), Michael Osborne (Associate Professor in Machine Learning), Mariarosaria Taddeo (Researcher, Information and Computer Ethics), John Tasioulas (Professor of Politics, Philosophy and Law, King's College London), Greg Taylor (Research Fellow, Microeconomics of Internet mediated marketplaces), Taha Yasseri (Research Fellow, Computational Social Science). We willen hierbij graag opmerken dat noch het OII, noch een van de genoemde personen verantwoordelijkheid draagt voor de tekst in dit rapport.



Attribution-NonCommercial-ShareAlike 3.0

(CC BY-NC-SA 3.0)



1	De Frankensteinfactor in IT	5
2	Frankenstein: superintelligent en superdom	11
3	Frankenstein: angst voor onszelf	19
4	Frankenstein: angst voor einde van de cultuur	23
5	Frankenstein: angst voor de 'Big Four'	28
6	De FrankensteinFactor: CIO als therapeut	40
Bijlage A	Asilomar AI Principles (2017)	45
Bijlage B	Frankenstein Syndrome Questionnaire	47
Bronnen		49



Het is de ratio die de automatisering altijd heeft gedomineerd. Maar tegenwoordig lopen de gemoederen vaak hoog op als het over IT gaat. Bijna alles lijkt mogelijk nu kunstmatige intelligentie, ook wel Artificial Intelligence of AI genoemd, opnieuw ontwaakt uit haar winterslaap. Op steeds meer terreinen zien we computers dingen doen die tot nog toe aan mensen voorbehouden leken. Dat is een fascinerend gegeven. Computers die luisteren, leren, kijken, praten – het belangrijke arsenaal van de kunstmatige intelligentie – komen steeds dichterbij in de buurt van het menselijk kunnen.

Wow!

Bijna dagelijks lezen we over een nieuwe doorbraak op het gebied van AI. Na schaken en het spel Go tonen algoritmes zich nu ook in het pokerspel superieur. Het onderzoek naar kanker ondergaat een enorme impuls en richt zich met AI 'Moonshots' op baanbrekende nieuwe behandelmethodes.¹ De ambitie van 's werelds grootste hedge fund, Bridgewater Associates, illustreert dat de grenzen van de AI-mogelijkheden steeds verder opgerekt worden. Bridgewater wil 75 procent van alle managementbeslissingen (tot en met het aannemen en ontslaan van mensen) door een AI-toepassing laten nemen, en dat alles binnen 5 jaar. Chatbots zoals het Chinese Little Bing worden onze persoonlijke maatjes met wie we onze zielenroerselen delen. We zien Watson de juiste antwoorden geven op cryptische quizvragen in het Amerikaanse Jeopardy en de beste (menselijke) spelers verslaan en we zien auto's zelfstandig door steden manoeuvreren. En wat te denken van de kunstmatige artdirector AI-CD B

die voor de Japanse vestiging van reclamebureau McCann Erickson een script schreef voor een commercial voor kauwgum en de concurrentie aanbindt met de menselijke art director?²

We staan aan de vooravond van een bijzondere ontwikkeling waardoor het automatiseringshart sneller gaat kloppen. De mogelijkheden om betere, veiligere en vele malen effectievere systemen te bouwen die meer impact hebben, lijken onbeperkt. Maar er is nog iets anders aan de hand. Stop deze typisch menselijke vaardigheden in een algoritme en ons onderbewustzijn begint een rol te spelen. Onze emoties nemen de overhand zodra we menselijke eigenschappen waarnemen in niet-menselijke gedaanten (algoritmes). Dat biedt uitgelezen kansen om de band met de klant te verstevigen – lees hiervoor ons vorige rapport *Het bot-effect: friending your brand*. Het gaat dus om meer dan alleen die efficiency en effectiviteit waarmee de IT al jaren vertrouwd is. In *Het bot-effect*:

1 Zie bijvoorbeeld <http://www.digitaleng.news/de/nvidia-joins-cancer-moonshot-with-ai-platform-for-accelerating-research/>

2 Zie de beide commercials hier: <http://adage.com/article/creativity/check-commercial-mccann-japan-ai-creative-director/304320/>



De machine in concurrentie met de mens: ook creatieve beroepen, zoals het bedenken van commercials, worden overgenomen door algoritmes. Dergelijke 'man versus machine'-competities zien we steeds vaker.

friending your brand beschreven we het verschijnsel van antropomorfisme (uit het Grieks: *anthropos* is 'mens' en *morphe* is 'gedaante') en de binding met de klant die hierdoor versterkt wordt. Maar daartegenover staat onze angst, de angst voor het onbekende, de automata en de androïde verschijningsvormen van de techniek, en de angst ten langen leste het onderspit te delven. De paradox is dat we het genot van de AI-mogelijkheden willen proeven en zelfs dat we ons hechten aan de kunstmatige algoritmes. Tegelijkertijd is er een bindingsangst, het onheimische gevoel dat opspeelt als we geconfronteerd worden met onze digitale 'doppelgänger'.

Zo zijn er voorbeelden dat het eenvoudigweg plaatsen van een menselijke robot achter een receptie tot grote consternatie leidt. Een actie bedoeld als een enthousiasmerende gimmick – ga eens spelen met

een robot – wordt verkeerd opgevat. Vergelijkbaar is de situatie waarin het gebruik van een onschuldige plaatje van robots in een powerpointpresentatie op de een of andere manier als 'creepy' wordt opgevat waardoor de vergadering een heel andere kant opgaat dan bedoeld. Om het nog maar niet te hebben over de angst voor het verlies van de eigen baan, aangewakkerd door Facebookberichten met onheilspellende titels als 'Robots will steal your job' of 'Beware of a Robocalypse'.

De sociale acceptatie van AI

'Wie zich niet verdiept in het gevoelsleven van zijn klant, kent zijn klant niet', schrijft Forrester Research in een recent rapport.³ Uit onderzoek van de British Science Association blijkt dat ruim een derde van de Britse bevolking van mening is dat kunstmatige intelligentie op de lange termijn een bedreiging vormt voor de overleving

3 'If you don't understand their emotions, you don't understand your customer.'
Forrester Research, 'Understanding the role of emotion in customer experience', 2015,
<https://www.forrester.com/report/Understanding+The+Impact+Of+Emotion+On+Customer+Experience/-/E-RES122503>

van de mens. Bijna de helft van de respondenten (46 procent) zegt niets te zien in robots die op mensen lijken. Hetzelfde geldt voor robots met menselijke vaardigheden, emoties en/of een persoonlijkheid zoals in de films *Her* en *Ex Machina*. En 60 procent is van mening dat er dankzij de robots over 10 jaar minder banen zijn.⁴ In de top-90-lijst van America's Top Fears 2015⁵ staat 'Trusting AI to do Work' op plaats 32. 'Robots' staan op plaats 36, 'AI' op 39 en 'Tech I don't understand' op plaats 49. Deze AI-gerelateerde thema's staan allemaal boven zaken als angst voor naalden, schietincidenten, vliegen, kidnapping of vermoord worden door iemand die je kent.

Kunstmatig intelligente algoritmes maken, terecht of onterecht, angstige gevoelens bij mensen los; ze scoren hoog op de 'FrankensteinFactor', de angst voor de kunstmatige dubbelganger. We verkennen in dit rapport wat daarachter kan zitten. Niet om de angst op te poken of aan te jagen, maar simpelweg als ruggensteun voor een debat dat nu en de komende jaren verder gevoerd wordt. Ons appel is daarbij om behalve met de bekende efficiëntie en effectiviteit ook rekening te houden met die gevoelens van angst. In een formule verwoord: de drie E's, de Efficiëntie en Effectiviteit aangevuld met een derde, de Existentie, zijn gezamenlijk doorslaggevend voor het succes van het initiatief.

De vraag is hoe de ambitie om succes met

IT te realiseren zich precies voltrekt. De acceptatie van technologie is immers altijd een samenspel van de technologische mogelijkheid, economische haalbaarheid en sociale wenselijkheid. De fascinatie voor een technisch hoogstandje ('Kijk! Gaaf! Een zelfrijdende auto!') betekent nog niet automatisch dat iedereen er een wil bezitten, ook al zou deze betaalbaar zijn. Is die auto wel te vertrouwen? Hoe ziet het algoritme eruit? Heb ik zelf nog de controle of ben ik eraan overgeleverd? Zomaar een paar vragen, sommige van meer filosofische aard, over transparantie, vertrouwen en zaken die te maken hebben met onze existentie.

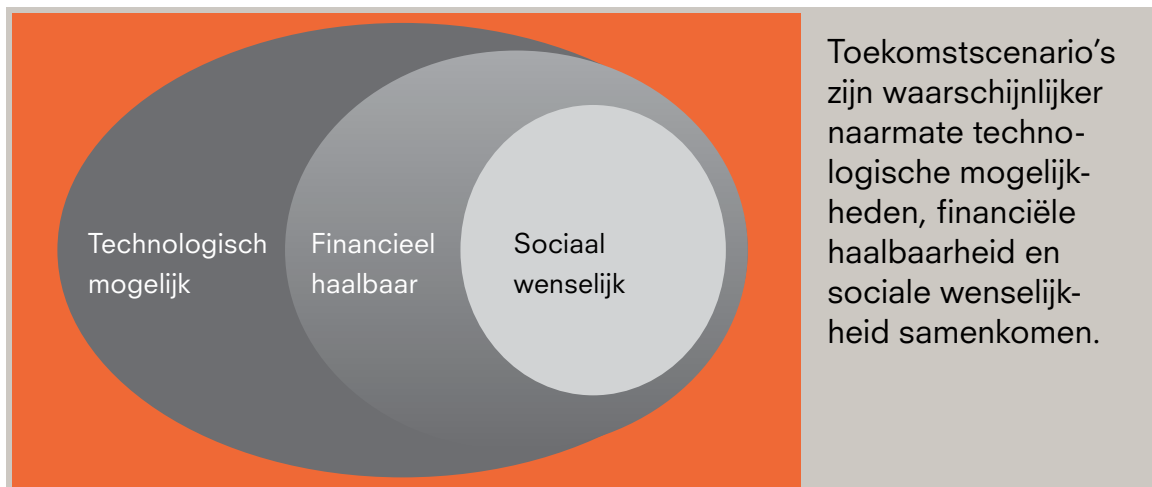
De vraag wat AI met het leven van uw klanten en uw medewerkers doet, hun bestaansrecht, hun identiteit, de controle die ze over hun zaken hebben, hun *raison d'être*, spelen veel vaker op naarmate er meer AI-toepassingen op de markt komen (en besproken worden). Dat maakt dat diegenen die verantwoordelijk zijn voor het IT-beleid in organisaties, directies en hun CIO offices, betere gesprekspartners moeten worden op dit vlak. Kennis en begrip van het gevoelsleven worden simpelweg belangrijker naarmate de technologie gebieden binnentreedt die dit gevoelsleven aangaan.

In dit rapport

Dit rapport vormt in zekere zin een tweeluik over binding en over bindingsangst voor AI. Het is een tweeluik samen met

Efficiëntie × Effectiviteit × Existentie = IT-succes

4 'One in three believe that the rise of artificial intelligence is a threat to humanity', <http://www.britishecienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity>
5 <https://blogs.chapman.edu/wilkinson/2015/10/13/americas-top-fears-2015/>



ons voorgaande rapport *Het bot-effect: friending your brand*. Daarbij gaat het in dit rapport niet zozeer over de vraag of deze angsten terecht zijn of niet. We verdiepen ons in het waarom van die angsten en waar zij vandaan komen. Het signaal dat we willen geven is dat er belangrijke winst te behalen is als de gevoelsrelaties meer worden meegenomen in de AI-plannen. Iedereen die sociale acceptatie van AI wil bereiken, doet er goed aan de anatomie van de angst in beschouwing te nemen. Iedereen die de mogelijkheden van deze geavanceerde technologie wil verkennen, zou de existentie centraal moeten stellen in de benadering.

Vier invalshoeken

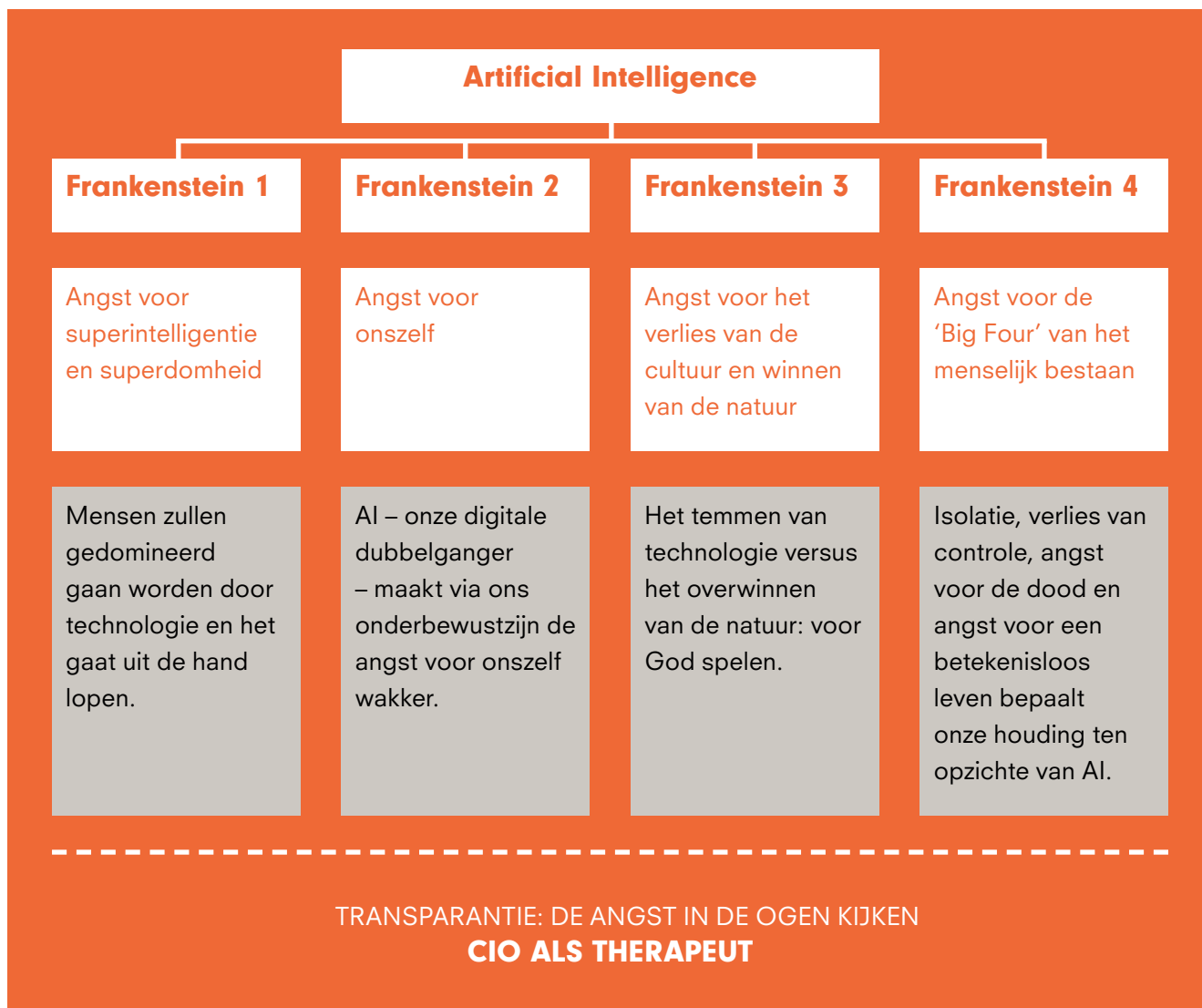
Voor de anatomie van de angst presenteren we vier verschillende invalshoeken die een kijk geven op hoe de relatie tussen technologie en emotie eruitziet. We beginnen onze beschouwing met een blik op het huidige debat: de angst voor superintelligentie, maar ook de risico's en

angsten die ontstaan als de systemen te dom blijken te zijn. De angst voor onszelf, zoals Freud die onder andere beschreef in 'Das Unheimliche', en de (Big Four) existentiële angsten van de neofreudische stroming in de psychologie geven vervolgens een dieper inzicht in deze onderbewuste gevoelsrelaties. Een vierde anatomische verkenning gaat over de angst voor de teloorgang van de cultuur, een argument dat ook een prominente rol speelt in een nieuwe Europese resolutie waar we later op terugkomen. Bij dit punt kijken we naar de verschillen tussen Oost en West, en dan met name naar Japan, en beginnen we met de vraag waarom de griezelfactor in Japan een kleinere rol lijkt te spelen.

Ten slotte zetten we de CIO op de stoel van de therapeut. Voorheen werden hem al vele functies toegedicht, van Chief Information Officer en Innovation Officer tot Change Officer. Klantentherapeut of kenner van het gevoelsleven, en dan met name de donkere kant van de angst,

behoort standaard nog niet tot het repertoire van de CIO. Maar in de wetenschap dat de angsten die we eerder beschreven vroeg of laat opduiken bij ieder AI-initiatief, lijkt het verstandig kennis op dit nieuwe terrein te verwerven. De rol van de therapeut hierin is driedelig: a) transparant zijn over de werking van de AI-algoritmes, over de plannen die er liggen op dit

terrein en over de te verwachten impact; b) de dialoog aangaan met gebruikers, klanten en medewerkers over de mogelijke angsten; en c) de lessen die u uit deze therapeutische dialoog trekt direct toepassen op uw eigen organisatie.



Frankenstein 1: het actuele debat

De angst voor superintelligentie en superdomheid

Vooraanstaande wetenschappers waarschuwen voor het einde der tijden als computers intelligenter worden dan de mens. AI-software-engineers waarschuwen juist voor de effecten als AI-toepassingen nog niet slim genoeg zijn. Het is een intellectueel welles-nietesdebat en Hollywood gaat daar dwars doorheen door in te spelen op het onderbuikgevoel van mensen.

Frankenstein 2: Freud en 'das Unheimliche'

We vrezen niet de robots maar onszelf

Psychoanalytici Sigmund Freud en Ernst Jentsch geven een verklaring voor het griezelgevoel dat mensen ervaren als ze met automata geconfronteerd worden. Hier gaat het niet om een cognitieve en bewust beredeneerde angst zoals in het actuele debat, maar juist om de relatie met die onderbuik, ons onderbewuste.

Frankenstein 3: het Oosten versus het Westen

We vrezen de verandering van de cultuur

Vanuit cultureel perspectief bestaat er een belangrijk verschil tussen Oost en West in relatie tot de angst voor AI. Vaak horen we dat men in Japan helemaal niet bang is voor robots en deze als 'vriend', niet als 'vijand' beschouwt. Het Japanse temmen van de technologie staat tegenover het overwinnen van de natuur in de westerse cultuur. Het is het contrast tussen de shintoreligie en de periode van de verlichting en de romantiek. We zullen echter zien dat beide culturele routes – Oost en West – op verschillende manieren leiden tot angsten voor kunstmatige intelligentie.

Frankenstein 4: de angst voor de Big Four

We vrezen allemaal hetzelfde

De vier aangeboren basisangsten waar ieder mens mee te maken heeft, schijnen een nieuw licht op de zaak. De angst voor kunstmatige intelligentie wordt erdoor aangewakkerd, maar tegelijkertijd kan AI ook een dempende werking hebben op deze angsten. Dit contrast roept nieuwe ethische vragen op. Is het doel om met AI deze angsten te reduceren nastrevenswaardig? Deze anatomie is gebaseerd op onderzoek uit de experimentele existentiële psychologie (XXP) en de terror-managementtheorie (TMT).

FRANKENSTEIN: SUPERINTELLIGENT EN SUPERDOM

'As the fate of the gorillas now depends more on humans than on the species itself, so would the fate of humankind depend on the actions of the machine superintelligence.'⁶

Nick Bostrom, filosoof en auteur van het boek *Superintelligence*

Precies tweehonderd jaar geleden begon de toen achttienjarige Mary Shelley met het schrijven van de klassieker *Frankenstein; or The Modern Prometheus*. In dit macabere verhaal wekt de excentrieke wetenschapper Victor Frankenstein dood materiaal tot leven. Victor probeert een vriend en metgezel te scheppen uit lichaamsdelen die afkomstig zijn van diverse lijken geroofd van het plaatselijke kerkhof. Zijn schepping vervult hem echter met afschuw en in paniek ontvlucht Victor zijn laboratorium. Hij laat zijn mislukte creatie achter in eenzaamheid, angst, verdriet en verwarring. Uit wraak vernietigt het monster alles waar zijn maker zo hartstochtelijk van houdt.

Angst voor Frankenstein en aanverwante figuren is een terugkerend thema geworden van sciencefiction-

boeken en -films. Het is een wereld van mutanten en bangmakerij, van halfgoden en fabelfiguren, van monsters van andere planeten en van killer robots. Isaac Asimov, een andere bekende sciencefictionschrijver, muntte dit angstfenomeen als het 'Frankensteincomplex': op mensen lijkende entiteiten (androïden) of mechanische versies daarvan (automata) jagen ons angst aan. Meestal lopen deze verhalen niet goed af voor de mens en is extinctie van de soort de uitkomst.

De fascinatie van de mens met zijn creaties, met de bedreiging van de mensheid door iets dat groter is dan hijzelf, de fascinatie met de dood en met overleven of zelfs onsterfelijk te worden, is onlosmakelijk verbonden met ons mens-zijn en onze geschie-

denis. In de dertiende eeuw bouwde Albertus Magnus, een katholieke bisschop, naar verluidt een androïde verschijning die kleine huishoudelijke taken vervulde.⁷ De bekende filosoof en theoloog Thomas van Aquino, leerling van Magnus, vernietigde de androïde, omdat hij er te veel door in verwarring werd gebracht. Er gaan verhalen dat de androïde van vlees en bloed was en dat Aquino gek werd van het oeverloze gezwam van de 'robot'. Ook de Griekse mythologie geeft ons een voorganger van de ijzeren man, Thalos, een bronzen automaton vervaardigd door Zeus om Europa te verdedigen tegen de ondergang. Een moderne variant van deze ijzeren man komt terug in de bekende Iron Man-films. De films zijn gebaseerd op de stripheld Tony Stark (voor het eerst in stripvorm verschenen in 1963) en de eerste film uit 2008 bracht in de eerste week 98.618.668 dollar op. Eén ding moeten we Hollywood en de entertainmentindustrie nageven: ze weten als geen ander munt te slaan uit kunstmatige intelligentie.

Hollywoodhorror, game over

Zonder verdere reflectie concluderen we al snel: kunstmatige intelligentie komt er onvermijdelijk aan, dus 'game over' voor

de mens.⁸ Omdat het brein zo belangrijk is voor onze identiteit, wordt AI al snel opgevat als de dief van onze eigenheid, zo zegt Kevin Kelly. Hier ligt de oorzaak van alle problemen en angsten. Kelly is de oprichter van het fameuze *Wired Magazine* dat de ontwikkelingen van de digitale cultuur op de voet volgt. Met zijn recente boek *The Inevitable* levert Kelly, een verklaard techno-optimist, zijn bijdrage aan het levendige AI-debat. In zijn boek stelt Kelly dat de gehele mensheid erop vooruitgaat en dat AI uiteindelijk goed zal zijn voor iedereen. Maar Kelly twijfelt of zijn boodschap overkomt. Het is erg lastig om op te boksen tegen de Hollywood-scenario's.

'There's a lot of possibilities, while in the alternative Hollywood version, there's only one story: we die.'



Hollywoodhorror: films als *Metropolis*, *Frankenstein*, *Terminator*, *Ex Machina* en *Her* presenteren al decennialang angstaanjagende toekomstbeelden

7 Godwin, W., *Lives of the Necromancers or The Exercise of Magical Power*, 1876, p. 180, <https://archive.org/details/livesnecromance04godwgoog>

8 'Wired founder Kevin Kelly on letting go of AI anxiety', <https://slackhq.com/wired-founder-kevin-kelly-on-letting-go-of-ai-anxiety-bcd94e50dedc#.amf4qbyxp>

Volgens Kelly is het idee dat wijzelf niet meer relevant zijn, niet meer nodig zijn, de bron van alle angst voor AI.

Hiermee snijdt Kelly een belangrijke vraag aan: waarom zijn we eigenlijk zo bang voor kunstmatige intelligentie? In een studie van Stanford University, de 'One Hundred Year Study on Artificial Intelligence' (ja, u leest het goed, de studie duurt honderd jaar), constateren de onderzoekers in hun eerste rapport uit 2016 bij het grotere publiek een overdreven angst, maar ook een overdreven optimisme. Zowel de optimisten als de angsthazen worden bestempeld als 'onrealistisch'. De stap van slimmere algoritmes en slimmere producten – van koelkast, auto of zorgrobot – naar een Hollywoodscenario is immers gauw gemaakt. Althans in ieder geval in gedachte, waarschijnlijk gedreven door ons onderbewustzijn. De o zo realistisch lijkende scenario's voor de introductie van uw nieuwe 'slimme' product of dienst stranden zomaar op die 'onrealistische' gevoelens van de beoogde klant.

Tallose gerenommeerde wetenschappers, kranten, websites en CEO's van technologiebedrijven waarschuwen voor de risico's van dit nieuwe tijdperk. Hollywood doet er

nog een schepje bovenop en schotelt ons voor hoe het zal aflopen, en dat is meestal niet goed. De optelsom van het mediage-weld is dat we op ons *qui-vive* zijn als de term 'kunstmatige intelligentie' valt.

Bemoeienis van overheden in Europa en Verenigde Staten

De angst lijkt er in Europa goed in te zitten. Een conceptresolutie die eind mei 2016 naar het Europese parlement is gestuurd en in februari 2017 is aangenomen, beschrijft de vrees dat kunstmatige intelligentie de intrinsieke Europese en humanistische waarden aantast.⁹ In de eerste paragraaf wordt Mary Shelleys Frankenstein naadloos verbonden met de ontketende opkomst van robots en slimme algoritmes. De resolutie verwoordt de bestaande risico's en angsten over de ontwikkeling van kunstmatige intelligentie en de acties die ondernomen moeten worden.

Het stuk spreekt onder andere over de-humanisering wanneer de robots in de gezondheidszorg menselijke taken overnemen en stelt voor dat robot-engineers een *code of conduct* tekenen die de 'dignity, privacy and safety' van mensen centraal stelt.

Een IEEE-rapport van december 2016, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, schetst een redelijk gedetailleerd beeld van hoe zo'n mogelijk verplichte gedragscode (waar de Europese resolutie over spreekt) eruit zou zien. Zo pleit het IEEE



9 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN>

bijvoorbeeld voor AI-systemen die transparant moeten zijn in hun voorspellingen. Dat betekent onder andere dat ze achteraf moeten kunnen 'uitleggen' wat ze gedaan hebben.

Het rapport voor de Europese Commissie pleit voor een Europees agentschap voor robotica en AI dat standaarden moet creëren en maatregelen moet nemen ter bescherming van de consument. Alle 'smart robots' dienen geregistreerd te worden bij dit instituut. Het is een van de vele voorstellen om ervoor te zorgen dat er veel meer transparantie komt in de algoritmes en de werking van de (ro)bots.

In het Verenigd Koninkrijk is een vergelijkbaar rapport gepubliceerd: *Robotics and artificial intelligence*.¹⁰ Een van de actiepunten uit dit regeringsstuk is om maatschappelijke discussie over AI op te poken en hiertoe een speciale commissie in het leven te roepen. Het Verenigd Koninkrijk sluit zich verder aan bij de strenge privacy-wetgeving van Europa.

De Amerikanen hebben hun 'National Artificial Intelligence Research and Development Strategic Plan' van de National Science and Technology Council (NSTC). Deze beleidsnotitie adviseerde president Obama in oktober 2016 over de impact van de 'transformative technology', zoals kunstmatige intelligentie wordt genoemd. Hierin lezen we dat AI-agenten vooral 'goed gedrag' moeten vertonen, net zoals je dat mag verwachten van je medemens. Ook de Amerikanen hameren op transparantie. Mensen moeten makkelijk kunnen

begrijpen wat de AI-toepassing doet en op welke informatie deze haar beslissingen baseert. Amerika en Europa pleiten dus beide voor transparantie en goed gedrag. Maar de Amerikanen verwoorden ook nog een andere angst: de angst om de economische wedloop van de Chinezen te verliezen als het om AI gaat en daardoor niet alle potentie te benutten.

Rem op innovatie

U kunt alle genoemde stukken downloaden en lezen via de links die in dit rapport zijn opgenomen. Een lijst van 23 AI-principes in bijlage A van dit rapport geeft een aardige kijk op de issues die in deze rapporten aan de orde worden gesteld. Deze 23 regels worden de 'Asilomar AI Principles' genoemd, verwijzend naar de plaats waar begin 2017 een AI-conferentie plaatsvond. Academics, AI-deskundigen en vermaarde kopstukken als Stephen Hawking en Elon Musk gebruikten de genoemde regeringsstukken als input voor het formuleren van hanteerbare principes voor onze AI-toekomst. Regulering vanuit overheden, de industrie en academia kan zeer wel een rem op innovatie veroorzaken. De vier rapporten (van de Amerikaanse overheid, het Europees Parlement, het Britse House of Commons en het IEEE) pleiten nu allemaal nog voor maatregelen om AI veel transparanter te maken. Na transparantie en inzichtelijkheid wordt het eenvoudiger om aanvullende regels op te stellen voor het door de Amerikanen bepleite 'good behaviour' of voor de bescherming van 'dignity, privacy and safety' waar Europa en het Verenigd Koninkrijk zich hard voor maken.

10 House of Commons Science and Technology Committee, *Robotics and artificial intelligence*, september 2016, <https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

Het loopt uit de hand: superintelligentie

Het opzienbarende begin van de Europese resolutie, de verwijzing naar het verhaal van Frankenstein, is relevant in het huidige debat over de angsten die met de opkomst van AI gepaard gaan. Frankenstein is immers maar een verhaal, een fantasie, en een nuchtere beschouwing over het gevaar voor AI wordt vertroebeld door zo'n sterke verhaallijn. Althans, dat is de overtuiging van de school die graag nuchter naar de feiten blijft kijken; we noemen dit 'kamp 2'. 'Kamp 1' daarentegen wordt vertegenwoordigd door mensen die waarschuwen voor superintelligentie en daarbij zelfs de hyperbolen van Frankenstein niet schuwen om ons te overtuigen van het gevaar. In de dagelijkse praktijk echter hebben we niet veel met de robots, mutanten en griezelfiguren uit Hollywood te maken. De snelle opkomst van

machine-intelligentie zorgt er wel voor dat die griezelfactor toeneemt. De grenzen tussen mens en machine vervagen immers. Het belangrijkste 'nieuws' dat de media domineert, is de mogelijke komst van superintelligentie waardoor alles uit de hand loopt. Nick Bostrom, filosoof en auteur van het boek *Superintelligence*, komt hiervoor speciaal met een eigen *code of conduct*, de MAXIPOK-regel: Maximale (MAXI) kans (Probability) op een goede uitkomst (OK) van het algoritme inbakken, want een kleine afwijking hiervan zou op den duur tot heel ernstige gevolgen kunnen leiden.¹¹ Bostrom verwijt overigens de entertainmentindustrie dat die goedkoop wil scoren, bijvoorbeeld door van artikelen over AI 'clickbait' te maken. Maar en passant jaagt hij ons flinke angst aan met uitspraken zoals dat we kleine kinderen zijn die spelen met een bom en dat AI gevaarlijker is dan de opwarming van de aarde.

Kamp 1: Het einde van de mensheid is nabij

Elon Musk (CEO van Tesla), Steve Wozniak (medeoprichter van Apple), Bill Gates (oprichter van Microsoft), Stephen Hawking (fysicus) en de eerdergenoemde Nick Bostrom behoren tot het kamp dat zegt dat we uiteindelijk de race tegen de computer verliezen en dat dit uiterst gevaarlijke situaties oplevert. Bostrom denkt dat het best nog wel vijf eeuwen kan duren, maar misschien duurt het slechts vijftig jaar. Gates verklaart niet te begrijpen waarom niet veel meer mensen zich druk maken over superintelligentie. Wozniak vermoedt dat de dominante computers ons zullen behandelen als hun huisdieren en beschouwt dit als goed nieuws omdat mensen in het algemeen immers goed voor hun huisdieren zorgen. Musk spreekt over het dreigende einde der mensheid en gelooft in de (minieme) kans dat ons leven nu al onderdeel is van een geavanceerde computersimulatie. Door deze zorgen gedreven investeert Musk significant (met andere investeerders 1 miljard dollar) in Open-AI omdat hij de technologie wil 'democratiseren' en wil voorkomen dat deze exclusief in handen komt van enkele grote spelers. Shane Legg, van het door Google opgekochte bedrijf DeepMind dat recentelijk de wereldkampioen versloeg in een spelletje Go, is ook bang voor de toekomst. Dit in tegenspraak met zijn baas Eric Schmidt, de topman van Google, die dergelijke angsten allemaal maar onzin vindt.

11 <https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>

Kamp 2: Geraaskal. Er zijn belangrijkere zaken waar we ons druk over moeten maken.

Er zijn gelukkig ook mensen die zich minder zorgen maken over een aanstaande computer-suprematie. We noemen Eric Schmidt (executive chairman van Alphabet), Pedro Domingos (machine intelligence professor aan de Universiteit van Washington en auteur van *The Master Algorithm*) en Luciano Floridi (filosoof, professor aan de Universiteit van Oxford en auteur van *The Fourth Revolution*). Domingos komt met een ontnuchterende conclusie: het echte probleem is niet dat de machines te slim worden en dat daardoor alles uit de hand loopt, maar dat de mensen te dom zijn om te beseffen dat we dit punt allang zijn gepasseerd. Leg het internet plat en we stevenen onvermijdelijk af op een catastrofaal einde van onze beschaving. Schmidt vindt dat niet iedereen zich in het debat moet mengen, en dat we dat over moeten laten aan welingelichte software-engineers. Floridi ten slotte noemt het erg onwaarschijnlijk dat computers ooit superintelligent worden. De groeisprint die AI nu ondergaat, doet hij af met de vergelijking met het klimmen in een boom. Als je de top eenmaal hebt bereikt, betekent dit het einde van je reis en niet dat je op weg bent naar de maan. Deep Learning mag dan nu 'slim' genoeg zijn om zelfstandig beelden te herkennen, maar er is nog lang geen sprake van iets dat lijkt op autonoom bewustzijn.

Ook in militaire kringen vreest men voor kunstmatige slimheid. Een groot aantal vooraanstaande wetenschappers en technologie-experts is van mening dat we überhaupt niet aan autonome wapensystemen¹² moeten beginnen.¹³ Zij stellen dat, hoewel autonome wapens voordelen bieden, zoals het verminderen van het aantal slachtoffers aan beide zijden, deze niet opwegen tegen de risico's. Zo vrezen zij voor de gevolgen van het in verkeerde handen vallen van autonome wapens. Ook zouden landen sneller tot een militair conflict overgaan als het risico van slachtoffers aan eigen zijde kleiner is.¹⁴

Volgens rapporten van het Amerikaanse ministerie van Defensie spelen robots een steeds belangrijkere rol op het toekomstige slagveld.¹⁵ Vooruitlopend hierop besloten onderzoekers van Harvard Law School in de zomer van 2016 de term 'war algorithm' te munten.¹⁶ Het is een juridische beschrijving met als doel het zo lastige begrip 'autonomous' te omzeilen. Een algoritme voldoet aan de 'war algorithm'-classificatie als het 1) in computercode uitgedrukt wordt, 2) zelfstandig keuzes maakt en 3) in een gewapend conflict kan opereren, ook als het daar niet specifiek voor ontworpen is. Dat laat-

12 Autonome wapens worden ook wel Lethal Autonomous Weapons (LAW), Lethal Autonomous Robot (LAR) of simpelweg *killer robots* genoemd.

13 <http://futureoflife.org/open-letter-autonomous-weapons/>

14 Walsh, J. & M. Schulzke, 'The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?', United States Army War College, september 2015, <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA621793>

15 Stewart, J., 'Strong Artificial Intelligence and National Security: Operational and Strategic Implications', U.S. Joint Military Operations Department, 18 mei 2015, <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA622591>

16 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2832734

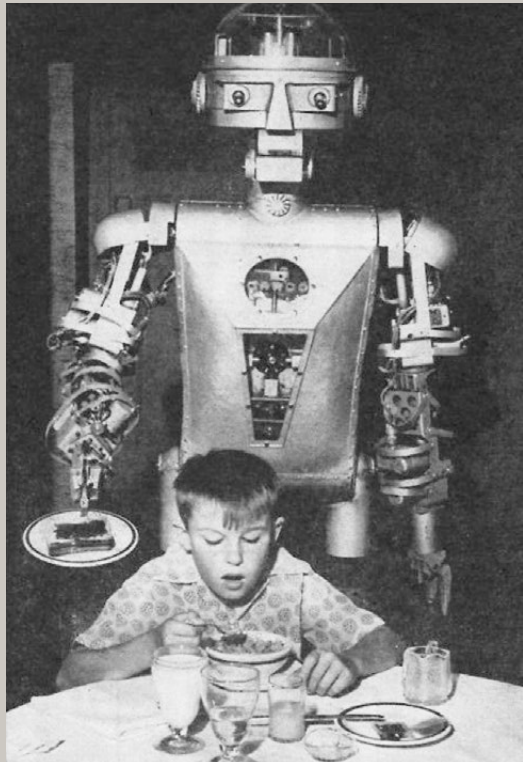
ste punt zou kunnen betekenen dat strikt genomen ook een zelfrijdende auto onder de juridische terminologie van oorlogs-algoritme valt.

Het loopt uit de hand: superdomheid

Tegenover de angst op lange termijn – superintelligentie – staat de angst op korte termijn voor de concrete problemen die worden veroorzaakt doordat AI op veel vlakken nog te dom is. Zoals de naam van het artikel al zegt: ‘Concrete Problems in AI Safety’¹⁷ gaat dit over veiligheidsproblemen die ontstaan als zelflerende systemen ontsporen. Het artikel verwoordt de zorgen van de AI-experts van Google en komt, gegeven de huidige beperkingen van zelflerende systemen, tot een opsomming van zaken waar we (echt) bang voor zouden moeten zijn. Om het idee te illustreren presenteren ze als voorbeeld de ‘ontbijtrobot’, een apparaat dat het ontbijt serveert en daarna de boel weer opruimt. Dit is wat er zoal mis kan gaan:

De ontbijtrobot en een aantal factoren waarvoor we bang zouden moeten zijn:

1. *Het veroorzaken van collateral damage.* De robot gooit vazen en andere spullen om, omdat hij zijn doel – het zo snel mogelijk serveren van het ontbijt – te strak najaagt.
2. *Het beloningshacken.* Als de robot zijn ogen sluit (de computer vision uitschakelt) ziet hij het ontbijt niet meer en hoeft hij dus niet meer af te ruimen. Het kan leiden tot ‘gaming the system’.
3. *Het missen van de nuance.* Tijdens het afruimen van de tafel gooit de
4. *Het opdoen van pijnlijke leerervaringen.* De robot moet experimenteren om te kunnen leren, bijvoorbeeld met kooktechnieken. Maar een ei in de magnetron en de onvermijdelijke explosie is geen wenselijke situatie.
5. *Het vertonen van onaangepast gedrag.* Een robot die geleerd heeft om ontbijt bij de burens te serveren en vervolgens bij u komt werken, kan veranderingen in de omgeving missen en fouten gaan maken.



De ontbijtrobot. Er kan van alles misgaan als hij niet slim genoeg is.

17 Amodei, D., e.a., ‘Concrete Problems in AI Safety’, Cornell University Library, 26 juli 2016, <https://arxiv.org/abs/1606.06565>

Aan dit lijstje van de experts van Google voegen we een drietal risico's toe die in het Global Risks Report 2017¹⁸ van het World Economic Forum worden benoemd:

6. *Cybercriminaliteit*. Risico's van het hacken van systemen in de fysieke infrastructuur, dat kan leiden tot een cascade van problemen.
7. *Een gebrek of een overdaad aan governance*. Te veel nieuwe wet- en regelgeving kan de ontwikkelingen afremmen en innovatie de nek omdraaien. Te weinig governance daarentegen laat de technologie mogelijk in een ongewenste richting ontwikkelen.
8. *Bugs in autonome wapensystemen* die tot catastrofale gevolgen kunnen leiden.

Hoewel je over deze kortetermijnrisico's minder hoort, zijn ze wel onderwerp van een actueel debat onder de engineers. Echter, toen Microsofts AI-chatbot Tay helemaal uit de bocht vloog, stonden de media er meteen bol van. Door doelbewuste pesterijtjes van het publiek, dat met

de chatbot in dialoog ging, werd Tay in een mum van tijd 'opgevoed' tot een racistische holocaustontkenner. Google moest al eerder zijn verontschuldigen aanbieden omdat Googles 'slimme' beeldherkenningsalgoritme aan foto's van donkere medemensen abusievelijk het toeschrift 'gorilla' toevoegde, met grote verontwaardiging tot gevolg. Dergelijke stigmatisering en belediging zijn pijnlijk, los van de vraag of ze met opzet (Tay) of per ongeluk (Google) gebeuren.

Vanuit de hoek van het cyberfeminisme maakt men zich erg druk over dit soort onbedoelde en vaak ondoordachte gevolgen. Erika Hayasaki, associate professor in literatuur en journalistiek aan de Universiteit van Californië, schreef in het blad *Foreign Policy* een uitgebreid betoog over de vraag of kunstmatige intelligentie goed is voor vrouwen of niet.¹⁹ Haar conclusie is dat seksisme, racisme en vrouwonvriendelijkheid op grote(re) schaal verspreid worden als we de slimme algoritmes niet meer regels opleggen. De door algoritmes aangeleerde vooroordelen van hun makers verspreiden zich gemakkelijk over het internet. Bovendien zijn de robots zelf vaak hyperbolen – sexy vrouwen, stoere mannen – die rolbevestigend in plaats van bevrijdend werken.



Tay, de AI-chatbot die ontspoorde.

18 World Economic Forum, *The Global Risks Report 2017*, <http://reports.weforum.org/global-risks-2017/executive-summary/>

19 Hayasaki, E., 'Is AI Sexist?' *Foreign Policy*, januari 2017, <http://foreignpolicy.com/2017/01/16/women-vs-the-machine/>

FRANKENSTEIN: ANGST VOOR ONSZELF

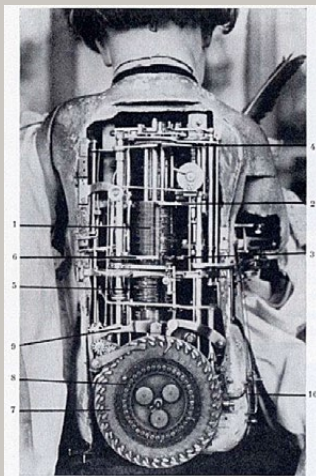
Laten we om te beginnen uit het voorafgaande concluderen dat kamp 1 zich extreem ongemakkelijk voelt bij het idee dat machines superintelligent worden, dat kamp 2 de gedachte bagatelliseert en dat de engineers nog wel even bezig zijn om AI echt intelligent te maken. Naast deze beredeneerde angst, waarbij cognitieve inschattingen worden gemaakt over toekomstige gebeurtenissen, is er de onderbewuste angst die rechtstreeks vanuit de onderbuik komt. Floridi (kamp 2) heeft een remedie voor de angsten die niet beredeneerd zijn. Hij zegt dat we gewoon moeten stoppen met het onszelf inbeelden van angsten:

‘Stel je voor dat je een donkere kamer binnengaat in een onbekend gebouw. Je kan dan heel bang worden voor de monsters die mogelijk opduiken uit de duisternis, of je kunt gewoon het licht aandoen om te voorkomen dat je ergens tegenaan botst.’²⁰

Maar dat is makkelijker gezegd dan gedaan, want het onderbewustzijn, waar die angsten zich manifesteren, gaat per definitie zijn eigen weg. Sigmund Freud en zijn vakgenoot Ernst Jentsch schreven al honderd jaar geleden over de onderbewuste angsten voor de automata (we noemen dat hier voor het gemak even kunstmatige intelligentie). Freud schrijft in zijn artikel ‘Das Unheimliche’ (1919) dat de confrontatie met mensachtige verschijningen, zoals mechanische poppen, een unheimisch gevoel oproept, juist omdat we ons er zo mee verwant voelen. Wij herkennen in de mechanische poppen (of robots) onze dubbelgangers. Hij benadrukt dat het woord *heimisch* zowel ‘zich thuis voelen’, ‘zich verwant voelen’ betekent, als het tegenovergestelde: ‘geheim’ (*heimisch*, *heimlich*).

Grammaticaal heeft ‘unheimisch’ dus een dubbele bodem (niet-verwant en niet-geheim), wat past bij zijn theorie dat wij onszelf nauwelijks kennen en dus ook

In de analyse van Jentsch en Freud speelt Olimpia, de 'living doll' (die een mechanisme bevat om gebaren na te bootsen en geluiden te maken) uit het boek *The Sandman* van E. Hoffmann uit 1816 een belangrijke rol



niet transparant zijn voor onszelf. Dat is in deze context ook van belang, omdat de verwantschap die we voelen met de AI-dubbelganger ook een gevoel oproept dat te maken heeft met het niet goed kennen van onszelf. AI, of in dit geval automata, als spiegelbeeld confronteert ons ook met de 'slechte' kanten van ons karakter en daarmee met de mensheid als geheel.

Uncanny: het niet kennen van de ander

Freud spreekt in zijn artikel ook over de 'doppelgänger' en de onderbewuste confrontatie met de dood die altijd ergens onder de oppervlakte aanwezig is. Praktisch gezien komt het erop neer, zegt Freud, dat we over dit soort zaken als 'wilden' gaan denken. Het onderbewustzijn stuurt onze gedachten en gedragingen. Als we apparaten zien bewegen of mentale kunststukjes zien uithalen die je alleen zou verwachten van mensen, zegt

Ernst Jentsch in zijn 'On the Psychology of the Uncanny' uit 1906, dan vinden we dat griezelig. Freud ontleedt de griezelfactor in zijn stuk nog verder en constateert dat de angsten vooral ontspringen aan de verbeeldingswereld van boeken, het theater en de verhaaltjes voor het slapen gaan. Dat is volgens Jentsch ook de plek waar we die gevoelens willen ervaren en beleven. Maar fantasie en werkelijkheid lopen heel vaak door elkaar en er is niet veel voor nodig om de kinderlijke angsten (krokodil onder het bed of de doden die je komen halen) in mensen te wekken. Freud heeft het in zijn stuk in dit verband over de 'so-called' intellectuele mensen die toch weer altijd door hun niet-intellectuele angsten gedreven worden. Het is interessant om te lezen in het stuk van Freud dat twee mechanismen die een 'uncanny' gevoel oproepen, nauw verwant zijn met de hedendaagse mogelijkheden van AI en Big Data: 'recurrent similarities' en de 'omnipotence of thought'.

Onverklaarbare toevalligheden

Een unheimisch gevoel ontstaat als we in toevallige gebeurtenissen een terugkerend patroon herkennen dat we niet kunnen verklaren. Een huis met nummer 62 voorbijlopen en een krantenkop lezen met het getal 62 precies op de dag dat je 62 bent geworden, is zo'n voorbeeld. In de wereld van Big Data waarin we vandaag de dag leven, ontdekken we dit soort 'onverklaarbare toevalligheden' steeds vaker. Freud zegt dat de herhaling van het getal 62 en de onverklaarbaarheid daarvan de interne angst oproept dat we een dwangneurose aan het ontwikkelen zijn en handelingen gaan herhalen zonder te weten waarom. Het argument waar Freud steeds weer op terugkomt is dat de angst voor de automata voortvloeit uit andere angsten die al ergens in ons verborgen zitten.

De wereld met je gedachten kunnen sturen

Een patiënt van Freud deed ooit de uitspraak dat een bepaalde persoon wat hem betreft beter dood kon neervallen. Toen dat een paar dagen later ook echt gebeurde, veroorzaakte dat bij die patiënt een enorme angst – een geloof dat gedachtesturing inderdaad mogelijk is. De angst van wat Freud de 'omnipotence of thoughts' noemt, zit nog steeds in ieder mens verscholen en ontwaakt door dit soort gebeurtenissen. In de AI-fantasie gaat het verhaal de ronde over de 'psychic pizza': alleen al het denken aan een pizza is voldoende om een pizzabezorger een half uur later bij je aan te laten bellen. Een grappige toekomstfantasie van AI-engineers kan zodoende in ons onderbewustzijn het gevoel aanwakken dat een alziend 'Evil Eye' echt bestaat.

IT waarbij je je vertrouwd voelt

Unheimisch. Maar zoals Freud al zei, het griezelige zit zowel in het bekende als in het onbekende. Het heimische van het AI-antropomorfisme, de mens als dubbeltanger waar men niet bang voor is, is een volledig nieuw terrein dat nu pas ontgonnen wordt. In ons vorige rapport *Het bot-effect: friending your brand* zinspeelden we al op de mogelijkheden van meer intieme relaties tussen organisaties en hun klanten. De (ro)bot als 'doppeltanger' van de organisatie is in staat andere emotionele connecties te maken. De grens tussen griezelig en vertrouwd,

intiem of ongewenst, lijkt zich te nestelen in het onderbewustzijn. Kennis van de subtiele verschillen die ervoor zorgen dat AI juist heimisch wordt, bepaalt in belangrijke mate het succes van de toepassingen in de markt.

Tot slot

Er is hernieuwde belangstelling voor theorieën die emotie en het onderbewustzijn centraal stellen. Hernieuwd zeggen we erbij, want na Freud is het met name de cognitieve psychologie geweest die de laatste decennia gedomineerd heeft: bijna al het menselijk gedrag kan verklaard

worden vanuit rationele afwegingen. Afwijkingen in het gedrag zijn volgens deze cognitieve school 'foutjes' in het menselijk systeem van 'informatieverwerking'. De moderne variant van Freud die sinds de jaren tachtig van de vorige eeuw in opkomst is, heet existentiële psychotherapie, oftewel experimentele existentiële psychologie (XXP). Die stroming ontkent niet het belang van Freuds werk, maar zegt dat er iets is dat nog net iets belangrijker is dan wat een mens in zijn jeugd heeft meegemaakt. Namelijk iets dat in ieders toekomst zeker gebeurt: de (onvermijdelijke) dood, de ontkenning en erkenning daarvan en hoe dat ons handelen bepaalt.

De dood en een aantal andere zaken die ons allen gegeven zijn, waarmee we altijd geconfronteerd worden, wie we ook zijn en waar we ook leven. Voordat we gaan kijken naar wat deze neofreudianen ons leren over angst voor kunstmatige intelligentie, zoomen we eerst in op de vraag waarom de griezelfactor in Japan minder lijkt voor te komen. De les die we daaruit kunnen trekken is dat Japan een lange traditie heeft in het temmen van technologie die de cultuur in stand houdt. In het Westen daarentegen is technologie sinds de verlichting veel meer in competitie met de menselijke natuur en lijkt ze de gedoodverfde winnaar.

Uncanny valley

Robotwetenschapper Masahiro Mori legt met zijn bekende 'uncanny valley' een relatie tussen de mate waarin een robot op een mens lijkt en de mate waarin we die robot griezelig vinden. Het centrale punt van zijn theorie is dat robots enger worden naarmate ze meer op mensen lijken (bijvoorbeeld door beweging), totdat ze weer zo veel op mensen lijken dat je het verschil niet meer ziet. Links ziet u robot C-3PO en R2-D2 uit *Star Wars*, Han van Hanson Robotics, daarnaast professor Hiroshi Ishiguro met zijn dubbelganger, Mark 1 van Ricky Ma, en rechts Peter Cushing die na zijn dood in 1994 recentelijk weer tot leven werd gewekt als levensechte computersimulatie in de *Star Wars*-film *Rogue One*. Overigens, hoewel plausibel, bestaat nog niet veel wetenschappelijk bewijs voor de werking van deze 'uncanny valley'.



FRANKENSTEIN: ANGST VOOR HET EINDE VAN DE CULTUUR

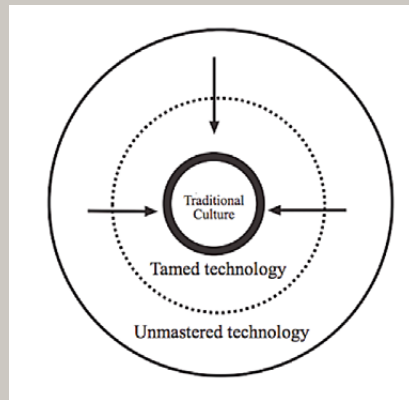
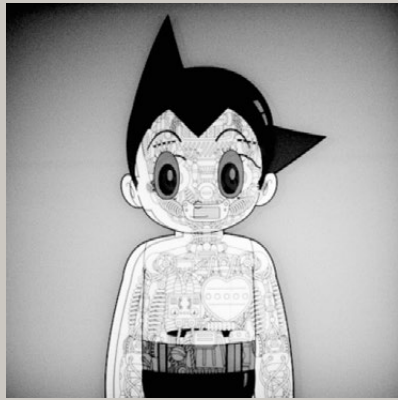
Hoe zit dat in Japan? Het is een veelgehoord argument, alleen in het Westen zijn we bang voor AI en robots. In het Oosten en met name in Japan ligt dat heel anders. De verklaring vanuit de verschillende opvattingen in de cultuur lijkt plausibel, maar de wetenschappelijke bewijslast voor een drastisch verschil in angstbeleving is beperkt.

Het standaardwerk van Frederic Kaplan, 'Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots', geeft een goed inzicht in de culturele verschillen en hoe die de reacties op menselijke robots verklaren. Kaplan werkte voor het Sony Computer Science Lab en geeft sinds enkele jaren leiding aan het Digital Humanities Lab van de Polytechnische Universiteit van Lausanne. Het lijkt, zo zegt Kaplan, of men in Japan geobsedeerd is door robots, omdat mensen er enthousiast op reageren en iedereen ze zonder al te veel problemen lijkt te accepteren. Kaplan draait de stelling echter om: juist wij in het Westen zijn geobsedeerd door robots en we accepteren ze daardoor minder snel. In Japan doet men er alles aan om de nieuwe technologie (robotica) naar de hand te zetten en ten dienste te laten zijn van de bestaande tradities. In het Westen lijkt het eerder op een gevecht.

De titel van het boek *Race against the Machine* van Erik Brynjolfsson en Andrew McAfee is een aardige illustratie van het sentiment dat het denken in het Westen overheerst. Het is óf de mens óf de

machine die wint. Kaplan gebruikt voor de Japanse aanpak het woord 'temmen'. Het temmen van de machines, de nieuwe technologie, is in de tweede helft van de negentiende eeuw zelfs formeel politiek beleid geworden. Japan leerde van de uit het Westen afkomstige technologische vernieuwing en zette die in om de traditionele Japanse cultuur te versterken in plaats van daarmee de strijd aan te binden. Het kunstmatige herschept het natuurlijke. Dat verklaart de hogere acceptatiegraad en mogelijk ook de Japanse aversie tegen cyborgs: half mens, half robot. Het integreren van de technologie in de cultuur staat centraal, niet het assimileren. Zodra de vreemde cultuur de overhand dreigt te krijgen, moet men er niets van hebben.

In de Japanse fictie, zoals stripverhalen, zien we terug hoe dit temmen van de technologie werkt. Bijvoorbeeld bij de populaire 'robot' Astro Boy (ook bekend als Mighty Atom), die rondwandelt met een atoomhart. Als zijn robotmaker hem in de steek laat, wordt hij door mensen vriendelijk opgenomen in hun samenleving. De robotverhalen in Japan gaan vaak over de



De getemde technologie ondersteunt de traditionele cultuur in Japan. Astro Boy (ofwel Mighty Atom) is een voorbeeld van hoe dat in de popcultuur werkt.

overwinning op iets dat van buiten komt, door iets te stelen van de technologie van de ander (atoomenergie) en daarmee de eigen cultuur te verdedigen. Dat Mighty Atom zo populair geworden is, nadat Japan de verschrikkingen heeft meegemaakt van de atoombom, onderstreept hoe ver deze drang tot integratie gaat.

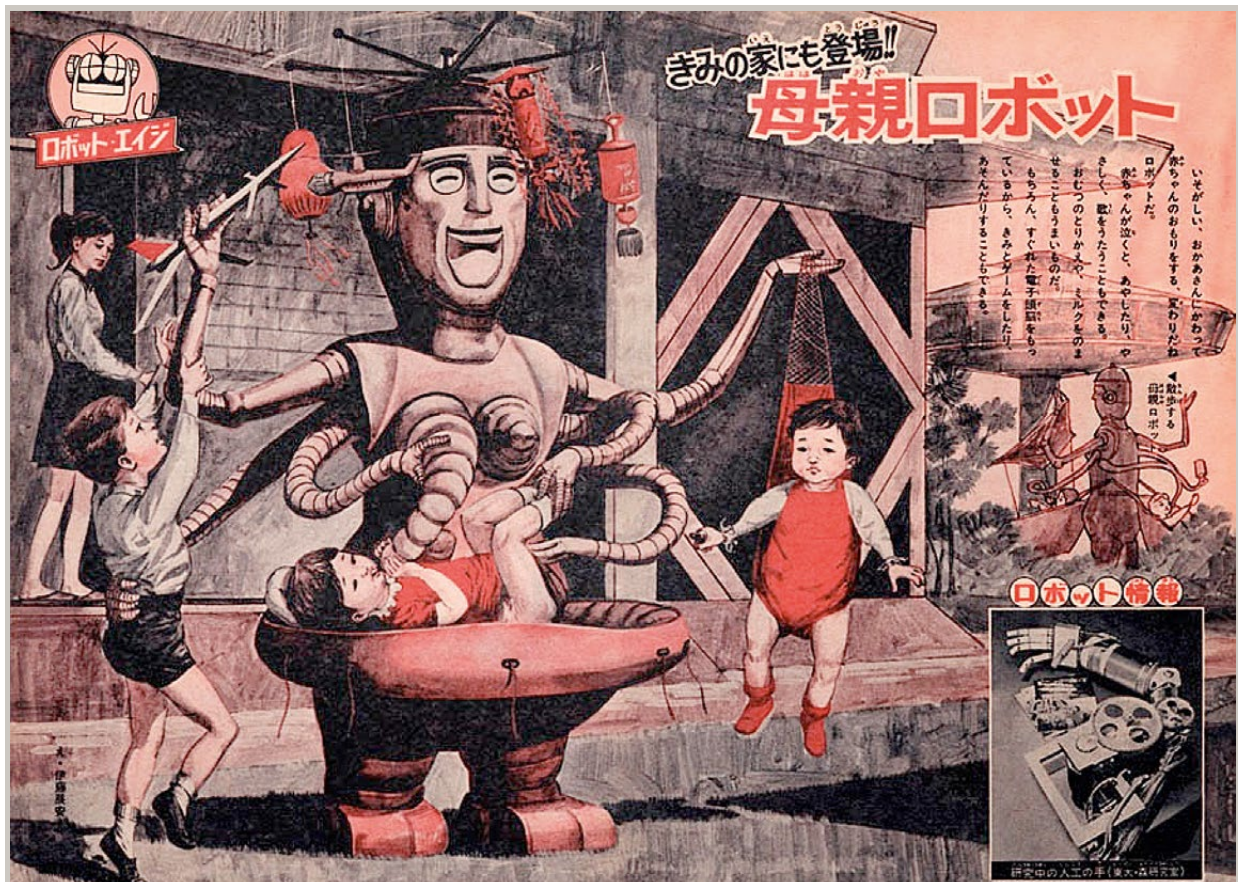
Deze eind-goed-al-goedverhaaltjes tref je overal in de Japanse robotfictie aan, inclusief verhalen over de megarobot waar je als kind in kan gaan zitten om superkrachten te verwerven. Het kind blijft in control en er is geen sprake van een autonome machine die de macht overneemt.

Zo groeien Japanse kinderen op met het beeld van goedaardige robots die je naar je hand kunt zetten en die vaak de rol van vader en moeder in zich verenigen. Die stripverhalen en films zijn een gevolg van de cultuur, maar tegelijkertijd ook een reflectie van het idee van getemde technolo-

gie waarmee de kinderen opgroeien. Kaplan beschouwt in hetzelfde artikel de relatie met robots in de westerse wereld. Hij betoogt dat de verlichting bij ons een cruciale rol heeft gespeeld in de manier waarop wij nu met robots omgaan. Tijdens de verlichting werden wetenschap, kennis en technologie verheven boven de natuur en dit creëerde de bekende tweestrijd: natuur versus cultuur (waar technologie en wetenschap een onderdeel van zijn). Het is in de westerse wereld vaak óf natuur óf cultuur, er is geen plaats voor hybride vormen. In Japan heeft men veel minder de behoefte de twee te scheiden. Daar worden natuur en het kunstmatige als een holistisch geheel beschouwd. Als voorbeeld schetst Kaplan het verschil tussen een karakteristieke Amerikaanse en een Japanse fontein. De Amerikaanse fontein spuit het water tientallen meters hoog, als triomf van techniek over natuur. De Japanse fontein is veel subtieler, een kabbelend watertje over een paar stenen,

precies zoals je het in de natuur verwacht aan te treffen, maar dan nog mooier. In de Japanse shintocultuur en -religie neemt het animisme een belangrijke plaats in. Dit is een spiritueel en filosofisch concept dat ervan uitgaat dat alles een ziel heeft. Planten, stenen, dingen, robots en mensen zijn volgens de shintofilosofie allemaal bezielde en daardoor ook meer met elkaar verbonden (en minder wezensvreemd).

Deze Japanse drang om de natuur te imiteren en te perfectioneren, versus de westerse drang om de natuur te overwinnen, is ook te verklaren vanuit die shintotraditie. Alles is immers een. Het illustreert het verschil tussen Oost en West. Het gaat in het Westen niet alleen om de scheiding van natuur en techniek, maar ook om de overwinning van de techniek op de natuur. In de westerse cultuur en filosofie wordt de ratio superieur geacht. Die ratio – nu



De mamabot. Japanse robot die zorgtaken uitvoert, speelt met de kinderen, luiers verschoont en de baby zoogt.

in een algoritmisch jasje van kunstmatige intelligentie gestoken – is vanzelfsprekend de bovenliggende partij die het van de menselijke natuur wint. In de westerse verhalen, de tegenhangers van de Japanse Atom-robot, loopt het dan ook meestal niet goed af voor de mens. De moraal – en de verhalen als reflectie daarvan – is dat de mens uiteindelijk niet competent genoeg is de natuur naar zijn hand te zetten. Je kunt beter niet voor God gaan spelen. Ook hier kan het argument van Freud opgevoerd worden dat de mens niet transparant is voor zichzelf en geneigd is slechte 'robots' te bouwen, omdat die slechtheid ook in ieder mens aanwezig is.

De romantiek en het werk van Rousseau, die volgden op de verlichting, brachten een verder spanningsveld teweeg. De opwaardering van de natuur uit die pe-

riode ging samen met de devaluatie van cultuur en techniek. Het kunstmatige drijft de mens af van de natuur, zo luidde de filosofie, en het geluk van de mens wordt aangetast. Kortom, de technologie en de automata die tijdens de verlichting nog als nobele producten werden gezien, kregen na de achttiende eeuw ook een negatieve connotatie. Dat spanningsveld komt ook tot uitdrukking in het eerder aangehaalde Frankensteinverhaal maar ook in het bijbehorende syndroom. Frankenstein wordt ook verlaten door zijn maker en niet, zoals in Japan het geval zou zijn geweest, liefdevol opgevangen door de mens. De superioriteit van het schepsel keert zich tegen ons en valt onze cultuur aan in plaats van daar onderdeel van te worden en ons te beschermen. De westerse moraal leert ons dat sollen met de natuur en ons erboven verheffen geen goed plan is.



Animisme leidt tot een begrafenis-ritueel voor een robot-hondje in Japan

Angst voor robots is angst voor onszelf en onze cultuur

De conclusie van Kaplan is dat de angst voor robots in feite de angst is voor het verliezen van wie wij zijn. Want de slimme apparaten kunnen ons in potentie dwingen tot het heroverwegen van datgene waarvan wij dachten dat het ons zo specifiek maakt. De angst voor robots is een reflectie van de angst om ons zelfbeeld ter discussie te stellen. Het is de angst dat de cultuur ten onder gaat, een angst die volgens filosoof Peter Sloterdijk gepareerd wordt door ons narcistische schild. Narcisme verheft en beschermt ons. Maar uiteindelijk moet de mens zichzelf opnieuw definiëren, gedwongen door de nieuwe machines met nieuwe capaciteiten die die van ons overtreffen.

De vragenlijst van Frankenstein

Of de Japanners ook daadwerkelijk minder bang zijn voor robots is wetenschappelijk nauwelijks aangetoond. Er zijn bijvoorbeeld studies die aantonen dat Amerikanen negatiever staan tegenover

stereotype robots dan Japanners, maar bij meer antropomorfe robots is dat precies andersom.²¹ Weer ander onderzoek toont aan dat jongens in het Verenigd Koninkrijk veel positiever zijn over robots dan jongens in Japan bijvoorbeeld.²² Angst en enthousiasme zijn universeel en de Oost-West-verschillen lijken kleiner dan de filosofieën ons doen geloven. Het meest gehanteerde onderzoeksinstrument om angstverschillen aan te tonen is de zogenaamde Frankenstein Syndrome Questionnaire. Het zijn dertig vragen die in onderzoek veelvuldig worden gebruikt om verschillen in opvatting in kaart te brengen. Omdat het een aardig inzicht geeft in hoe vanuit de wetenschap naar het complete angstpalet gekeken wordt, hebben we de hele lijst met vragen en stellingen in bijlage B opgenomen. Het is een breed palet dat varieert van de angst dat robots ons lui maken, de angst dat ze voor ons gaan oordelen en de angst dat ze onze kinderen gaan beïnvloeden, tot de angst dat we vergeten wat ons mensen nu eigenlijk tot mens maakt.

21 <https://www.bartneck.de/publications/2008/whoLikesAndroids/bartneckRoMan2008.pdf>

22 http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura_et al_AISB2015_differencesonsocial.pdf;sequence=2

FRANKENSTEIN: ANGST VOOR DE 'BIG FOUR'

'The more unlived your life, the greater your death anxiety. The more you fail to experience your life fully, the more you will fear death.'

Irvin D. Yalom, *Staring at the Sun: Overcoming the Terror of Death*

Op de vraag uit de voorgaande paragraaf of het goed of slecht is dat we de natuur naar onze hand zetten, kunnen we lang debatteren. De eventuele angsten die hiermee gepaard gaan, zijn in eerste instantie beredeneerde angsten. Dit in tegenstelling tot deze vierde en laatste component in onze anatomie: vier angsten waar we allemaal mee hebben te dealen zeggend. Deze zogenaamde 'Big Four' angsten uit de existentiële psychotherapie en de daaraan verwante scholen zijn altijd in meerdere of mindere mate aanwezig:

- de angst om geïsoleerd te raken van de wereld en van de mensen om ons heen;
- de angst om de controle te verliezen;
- de angst om een betekenisloos leven te leiden;
- de angst voor de dood.

Ook hier speelt het onderbewustzijn een grote rol en is het een neofreudiaanse kijk op de angst. Het meest markante verschil is dat Freud altijd achterwaarts kijkt op zoek naar verklaringen van het gedrag:

de geboorte, oerinstincten uit de evolutie, relatie met de ouders, traumatische ervaringen et cetera. De Big Four-theorie kijkt daarentegen met name vooruit, naar de onvermijdelijkheid van de dood waar we allemaal mee te maken hebben en dit existentiële gegeven dat een fundamentele verklaring geeft voor alles waar mensen bang voor zijn. De anatomie van deze Big Four angsten levert interessante nieuwe inzichten maar ook nieuwe vragen op. Twee willen we vooraf met name benadrukken, voordat we dieper ingaan op deze angsttheorie.

1. In de anatomie tot nu toe hebben we vooral negatief gekeken naar het angstverschijnsel. Maar met de Big Four als gegeven kan dat ook worden omgedraaid en kan AI gezien worden als iets dat onze existentiële angsten juist kan dempen.
2. Een tweede inzicht is dat de angst rond privacy niet tot de vier basisangsten behoort. Sterker nog, de angst om juist niet gezien te worden en diensengevolge een ongezien leven te

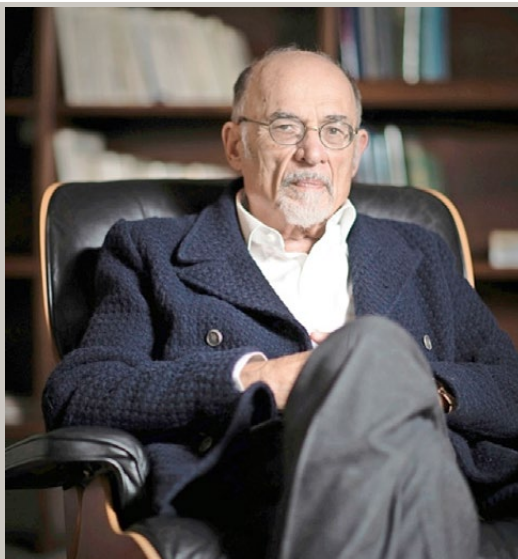
leiden, behoort wel tot de vier hoofdangsten. Dat is met name relevant in het AI-debat dat vaak gaat over privacyverlies, maar zelden over het tegengestelde: de kans om je te laten zien en gezien te worden. Mensen hebben nu eenmaal een 'need to be seen'.

Existentiële angst

Vragen over ons bestaan, over onze existentie, zijn dikwijls aanwezig, zij het op de achtergrond. Zo nu en dan wordt ons leven abrupt verstoord door het overlijden van een naaste, een ontslag, een midlife-crisis of de beëindiging van een relatie. Op dat moment worden wij ongenadig geconfronteerd met onze, voorheen sluimerende, existentiële angsten. Een mildere variant van een dergelijk groot persoonlijk verlies is bijvoorbeeld de angst die boven komt bij het binnenstappen van een vliegtuig, als we in de ogen kijken van een androïde of als we straks wegrijden in onze zelfrijdende auto.

De existentiële psychotherapie heeft een sterk filosofische achtergrond en is beïnvloed door werken van onder anderen Nietzsche, Kierkegaard, Husserl, Heidegger, Sartre, Maslow en de humanistische psychologie. Men spreekt ook wel over de experimentele existentiële psychologie (XEP) en de terror-managementtheorie (TMT) – allemaal scholen die nauw met elkaar verwant zijn en sinds de jaren tachtig van de vorige eeuw in opkomst zijn.

Een van de voornaamste hedendaagse experts van de existentiële psychotherapie is Irvin D. Yalom.²³ Yalom is al meer dan



Irvin Yalom, emeritus professor psychiatrie aan Stanford Universiteit.

zestig jaar in het vakgebied actief en heeft zich gespecialiseerd in groepstherapie en individuele existentiële psychotherapie. Zijn bijdragen aan dit beroepsveld, zoals de boeken *Existential Psychotherapy* en *The Theory and Practice of Group Psychotherapy*, hebben grote aantallen nieuwe psychologen en psychotherapeuten op weg geholpen. Ook met zijn fictieboeken waarin hij zijn theorieën op toegankelijke wijze vertaalt naar de levensverhalen van zijn personages, heeft hij een groot publiek geïnspireerd. Nietzsches *tranen* is misschien wel zijn bekendste non-fictieboek. In zijn boek *Existential Psychotherapy* beschrijft Yalom een van zijn meest invloedrijke bijdragen aan de existentiële psychotherapie, de vier 'ultimate concerns' waar elk mens mee te maken heeft.

23 Yalom pleit voor een inclusieve psychotherapie waarbij voor de oorsprong van psychologische problemen wordt gekeken naar de biologie en genetica (psyche-farmacologisch model), onze onderdrukte instincten (Freud), een disfunctionele relatie met ouders (relationeel perspectief), disfunctionele gedragspatronen (cognitieve gedragstherapie) en ook naar onze confrontatie met ons bestaan (existentiële benadering).

De vier concerns:

- **need to control**
(angst voor vrijheid);²⁴
- **need to be seen**
(angst voor isolatie);
- **need to survive**
(angst voor de dood);
- **need to be needed**
(angst voor betekenisloosheid).

Big Four: angst voor de dood

Het zijn vier afzonderlijke angsten, maar ze zijn onlosmakelijk met elkaar verbonden.²⁵ Ze gaan eigenlijk allemaal over de dood, of hoe we met deze ene zekerheid omgaan. Het gaat hier niet om de acute doodsangst als we oog in oog staan met een leeuw, maar om de angst voor de dood op de langere termijn en hoe we met de onvermijdelijkheid daarvan omgaan. Dat doen we bijvoorbeeld door er niet te veel over na te denken, maar wel lekker druk te zijn met belangrijke zaken waaraan wij betekenis ontleenen. Valt die betekenis weg, dan komt die existentiële angst weer boven. Alleen vanwege pedagogische redenen, zegt Yalom, is het handig dat we ze afzonderlijk bekijken.

Het handboek *Experimental Existential Psychology: Exploring the Human Confrontation with Reality* geeft een actueel overzicht van het huidige denken over deze vier angsten en de prominente rol die de dood (of het ontkennen daarvan) speelt. Daarin wordt ook verkend hoe onderwerpen als schoonheid, spiritualiteit

'The fear of death plays a major role in our internal experience; it haunts as nothing else does; it rumbles continuously under the surface; it is a dark, unsettling presence at the rim of consciousness.'

Irvin Yalom

en nostalgie in dit angstplaatje passen. Het terugkerende thema is dat het hier steeds om verwerkingsmechanismen gaat: omgaan met de angst voor de dood en het dempen van die angst die altijd ergens in het onderbewustzijn sluimert. De hang naar het verleden, melancholie en nostalgie bijvoorbeeld, geeft grip op het leven, creëert betekenis met het oog op de dood. Ook voor de drie andere angsten uit de Big Four leveren de 'need to be seen' (met uitwassen zoals het narcisme), de 'need to be in control' (met uitwassen naar de controlfreak) en de 'need to be needed' (de workaholic) eenzelfde soort compensatie.

24 Yalom ziet deze angst in relatie tot angst voor de ultieme vrijheid. Hoewel veel mensen vrij willen zijn, zegt Yalom dat de mens bang is voor ultieme vrijheid: het besef dat de werkelijkheid niet gestructureerd en objectief is. Wij moeten zelf deze structuur aanbrengen en dragen hierdoor verantwoordelijkheid voor onze eigen ervaringen in het leven.

25 <http://www.freepatentsonline.com/article/North-American-Journal-Psychology/173513562.html>

Big Four: angst voor een betekenisloos leven

De angst dat ons leven er niet toe doet, is een van de belangrijkste drijfveren in het leven: de zoektocht naar betekenis.²⁶ De angst dat het leven zinloos is, is volgens Yaloms collega, neuroloog en psychiater Viktor Frankl, zelfs de meest ultieme angst. De relatie met kunstmatige intelligentie ligt hier natuurlijk voor de hand. Kunstmatige intelligentie is in staat de taken over te nemen waar we betekenis aan ontleenen en daarmee de angstprikkels – betekenisloosheid – te activeren. Dit is ook het argument van Kevin Kelly uit het begin van dit rapport. Het snelle een-tweetje dat we maken: de robots komen, game over voor de mens. Zo vaak werd dit al in sciencefictionfilms belicht, maar de film *Terminator 2: Judgment*

Day verbeeldt dit misschien nog wel het beste. De moeder van de tienjarige John zegt, als ze de cyborg met haar zoontje ziet spelen, dat het haar opeens volstrekt helder is dat deze Terminator nooit moe zal raken. Hij zal altijd willen blijven spelen, altijd aandacht geven aan haar zoon, nooit dronken worden en hem slaan. Deze machine is de enige die de vaderrol echt waar kan maken.

Het gaat dus om de angst dat (alle) betekenis die ik aan het leven ontleen, mij wordt ontnomen door intelligente en affectieve apparaten. Mijn baan, mijn hobby en zelfs mijn relatie. We kunnen er lacherig over doen, maar ook op dat laatste vlak wordt er al volop gespeculeerd over de toekomstige mogelijkheden van seksrobots. Het is dus geen wonder dat de technologische vorderingen angstvallig – op dat niveau dus – worden beschouwd en beleefd.



Spelen met de gedachte dat robots de rol van mannen overnemen, bestaat al veel langer. Ook al zijn het maar simpele (en onvoorstelbare) scènes, onderhuids en onbewust kunnen ze de angst voor betekenisloosheid triggeren.

26 Frankl plaatst de wil tot zingeving boven Nietzsches wil tot macht en Freuds wil tot lust. Frankl, V., *Man's Search for Meaning*, Beacon Press, 1 juni 2006 (eerste publicatie: 1946).

Roy Baumeister, professor in de psychologie aan de Universiteit van Florida, beschrijft het mechanisme hoe zingeving leidt tot positieve eigenwaarde en een 'need to be needed'. Familie, vrienden, klanten, de organisatie en onze huisdieren hebben ons op een bepaalde manier nodig (vinden wij) en daaraan verbinden we vaak onze zin om te leven. We zien dat het mechanisme veel te maken heeft met onze wens de controle over ons leven in de hand te hebben. Voor zingeving moet u:

1. Een doel vinden, zodat u betekenis kunt geven aan uw activiteiten. U doet het ergens voor, zeg maar.
2. Vervolgens kunt u uw eigen activiteiten rechtvaardigen; u kunt uw gedrag verantwoorden.
3. Dat geeft het gevoel dat u een zekere grip of controle op uw leven heeft.
4. Een positieve eigenwaarde is het gevolg.

Het heikele punt: de banen

We hebben het hier niet zozeer over het maatschappelijke vraagstuk of massawerkloosheid dreigt als gevolg van AI, maar over het individuele geval van het verlies van *mijn* baan en *mijn* betekenis die ik daaraan ontleen. Dat laatste kan een heel concrete en realistische angst zijn op het moment dat mijn inkomen op het spel staat. Maar het verlies van betekenis kan daar niet los van gezien worden. De existentiële angst een zinloos leven te leiden of te hebben geleid nu de machine het werk overneemt, speelt net zo hard mee. Dan begint de zoektocht. Niet alleen naar een nieuwe baan, maar ook naar nieuwe betekenis. Aan beide angsten is

wat te doen: herplaatsing, bijscholing en bemiddeling op de arbeidsmarkt en therapeutische begeleiding voor het verlies van betekenis.

Kevin Kelly, de techno-optimist uit onze inleiding, beschrijft de rollercoaster waar mensen doorheen gaan op zoek naar die andere baan. Het is een psychologische les in zeven stappen waar maar één conclusie aan te verbinden is: inzichten kunnen veranderen, meningen en angsten zijn niet in beton gegoten. Deze Baumeister 2.0 versie²⁷ geeft mogelijk troost voor iedereen die robotisering op zijn carrièrepad krijgt.

The 7 stages of robot replacement

1. *A robot/computer can't possibly do what I do.*
2. *OK, it can do a lot, but it can't do everything I do.*
3. *OK, it can do everything I do, except it needs me when it breaks down, which is often.*
4. *OK, it operates without failure, but I need to train it for new tasks.*
5. *Whew, that was a job that no human was meant to do, but what about me?*
6. *My new job is more fun and pays more now that robots/computers are doing my old job.*
7. *I am so glad a robot can't possibly do what I do.*

27 Zie <http://kk.org/thetechnium/the-7-stages-of>



Het is diezelfde Kelly die ons waarschuwde voor de kracht van Hollywood. De horrorscenario's van de robots die komen en het einde van de mensheid inluiden. Zonder te pretenderen dat we weten of er wel of niet structureel meer banen verdwijnen dan er nieuwe bij zullen komen, willen we wel wijzen op het feit dat het angstscenario het een stuk beter doet in de media.

De onheilsverhalen over mechanisering en automatisering achtervolgen ons al vele decennia, zoals blijkt uit het bovenstaande krantenbericht uit 1928 en vele andere berichten van de afgelopen eeuw.

De onderstaande afbeelding illustreert de manier waarop het AI-nieuws zich via

sociale media verspreidt. De FrankensteinFactor doet het uitstekend als clickbait.

Er zijn tal van onderzoeken die wijzen op verlies van banen vanwege AI-automatisering.²⁸ De nuance dat technologie de afgelopen 140 jaar altijd voor meer banen heeft gezorgd, wordt er vaak wel bij vermeld, maar dat is niet wat blijft hangen. De OECD becijferde vorig jaar nog²⁹ dat het maar om 9 procent van de banen zou gaan die door AI 'weggeautomatiseerd' worden. Dit ligt in lijn met een recente studie van McKinsey³⁰ die zegt dat maximaal 5 procent van de banen volledig geautomatiseerd worden. We moeten volgens McKinsey veel meer rekening houden met een automatisering van onderdelen van huidige banen, dus specifieke taken in plaats van volledige banen.

Tot slot: de vierde revolutie

Het grotere plaatje begint zich af te tekenen. Wij (de mensheid) blijken steeds minder bijzonder dan we oorspronkelijk dachten. Oxford-professor Luciano Floridi legt hierop het accent in zijn boek *The Fourth Revolution*. Bij grote technolo-



28 Zie bijvoorbeeld: Frey, C. & A. Osborne, 'The Future of Deployment', 2013, http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Deloitte, 'Technology and people: The great job-creating machine', 2015, <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf>

29 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis', mei 2016, http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en

30 McKinsey Global Institute, 'Harnessing automation for a future that works', januari 2017, <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>

'Freedom in this sense has terrifying implications; it means that beneath us there is no ground – nothing, a void, and an abyss. A key existential dynamic then, is the clash between our confrontation with groundlessness and our wish for ground and structure.'³¹

Irvin Yalom

gische doorbraken en ontdekkingen zien we keer op keer hetzelfde patroon. De mens valt steeds een klein beetje meer van zijn sokkel. Dat zagen we tot nog toe vier keer gebeuren, de vierde revolutie is de IT-revolutie waar we nu middenin zitten. De eerste keer was toen we ontdekten dat de aarde niet het unieke stabiele centrum is van het totale heelal waaromheen alles draaide. Onze planeet beweegt net als andere planeten en draait om een ander centrum: de zon (Copernicus en Galileo). Eigenlijk niets bijzonders. De mens vormde niet langer het centrum van het universum. Dat we niet de unieke creaties zijn die we altijd dachten te zijn, maar net als alle dieren gewoon deel uitmaken van de evolutie (Darwin), is twee. En dat de mens zichzelf niet goed kent en niet zo transparant is als we dachten, maar dat we een duister gevoelsleven hebben dat we zelf niet eens goed begrijpen (Freud), is drie. We zijn niet de 'masters of the universe'. We zijn niet eens de masters van onze eigen geest, en nu ontdekken we dat we niet meer de al-

leenheersers zijn van de taal, het redeneren, het luisteren, het voorspellen en het analyseren. Dit is de vierde revolutie, die van de informatietechnologie, en die is al een tijdje bezig. Maar door de opkomst van AI worden we met de neus op de feiten gedrukt. Wat de technologie met ons doet, in dit geval kunstmatige intelligentie, is minstens zo belangrijk als wat wij met die technologie doen. Dit relativeren van onze eigen betekenis wakkert de angst voor de betekenisloosheid van het bestaan verder aan.

Big Four: angst voor controleverlies

De angst om de controle te verliezen is in meerdere of mindere mate bij iedereen aanwezig. In ieder mens schuilt wel ergens een controlfreak. De controlfreaks, ook wel de 'internals' genoemd, maken graag zelf afwegingen, verdiepen zich in de materie en houden graag het heft in

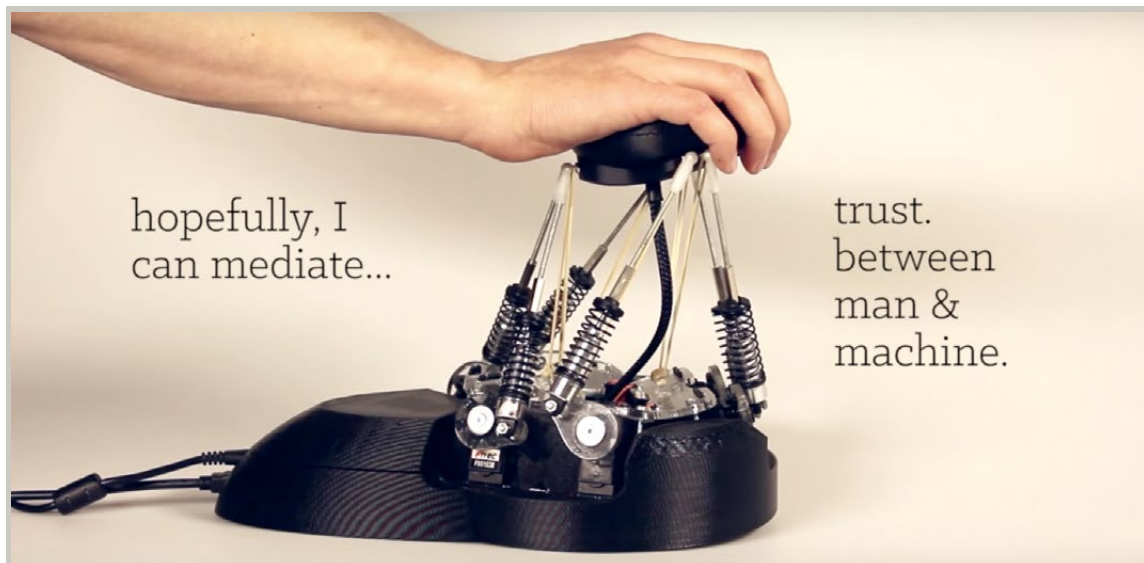
31 Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980.

32 Het gaat hier om het psychologische begrip 'locus of control'. Zie: Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980.

handen.³² De tegenhangers van de controlfreaks, de zogenaamde 'externals', leggen het lot graag in handen van anderen. Dat is ook een manier om ergens zelf geen verantwoordelijkheid voor te hoeven nemen. Zij nemen het eerder voor lief dat een chatbot of app voor ze beslist, zolang ze er zelf maar niet over na hoeven te denken. Maar dit zijn uitersten, een bepaalde mate van 'in control' willen zijn zit verborgen in ieder mens.

Veel mensen zitten bijvoorbeeld veel liever zelf achter het stuur dan op de bijrijdersstoel (en gaan zitten 'meerijden'), om over het stuur uit handen geven aan een algoritme dat de auto bestuurt nog maar niet

te spreken. Recente studies laten zien dat slechts een op de tien mensen volledig vertrouwen heeft in zelfrijdende auto's.³³ En 95 procent van de mensen wil het volledige controlepaneel (stuur, rem- en gaspedaal) behouden, ook al is de auto volledig in staat zelfstandig te rijden. Dat Mercedes-Benz zijn nieuwe zelfrijdende auto (de F 015) als 'quality time'-concept aan de man brengt is begrijpelijk, maar roept ook vragen op. Begrijpelijk, omdat de video-entertainment in de auto je tijdens de reis lekker laat genieten van mooie films (quality time). Maar de vraag is of het controleverlies en de angst die daarmee gepaard gaat ook wel echt quality time oplevert voor de inzittende. In



Stewart is een 3D joystick die als mediator een rol speelt tussen de zelfrijdende auto en de 'bestuurder'. De 'bestuurder' voelt via Stewart precies wat de zelfrijdende auto van plan is, bijvoorbeeld versnellend een bocht naar rechts nemen, en kan dit overrulen door Stewart een andere kant op te bewegen. Mogelijk een oplossing voor de mensen die moeten wennen aan het uit handen geven van controle.³⁴

33 <http://www.dailymail.co.uk/sciencetech/article-3245058/Do-trust-driverless-cars-1-10-feel-confident-getting-autonomous-vehicles-women-skeptical.html>

34 http://felixros.com/projects.html#big_wrap

ieder geval neemt Mercedes-Benz een van de angsten rond zelfrijdende auto's weg. Het bedrijf kondigde aan in alle gevallen te kiezen voor de veiligheid van de bestuurder. Het soort dilemma dat u vast al eens gehoord hebt: voor wie kiest de auto in geval van gevaar? Een moeder met een kindswagen, de bejaarde die de weg oversteeft ... et cetera. Het kan allemaal naar de prullenbak. De Mercedes kiest consequent voor de veiligheid van de inzittenden.

Autonomous of semi-autonomous

In feite worden we geconfronteerd met de angst voor de ultieme vrijheid, zo legt Yalom uit. Het gevoel geen enkele grond meer onder de voeten te voelen (figuurlijk) roept automatisch de reactie op dat we de touwtjes in handen willen houden om deze leegte op te vullen.

De les voor AI-toepassingen lijkt eenvoudig. Een bepaalde menselijke zeggenschap over autonome systemen heeft vanuit angstperspectief altijd de voorkeur. Het walhalla van AI, een wereld waarin alles door machines en algoritmes wordt gerund en de mens eenvoudigweg van zijn quality time kan genieten, staat op gespannen voet met deze angst die eenieder gegeven is. Als we niet zelf in actie komen om het leven naar de eigen hand te zetten en zodoende betekenis te geven aan het leven en daarmee onze angst voor de dood te verdringen, worden we juist bang en ervaren we het tegenovergestelde van quality time. Er zijn natuurlijk heel veel verschillende manieren om die zeggenschap over AI vorm te geven. Het hoeft geen belem-

mering te zijn voor de toepassing van AI, alleen dient rekening gehouden te worden met deze 'need to control'.

Big Four: angst voor isolatie

Het verhaal gaat dat toen Yalom werkte met mensen die hun partner hadden verloren, hij niet alleen was aangedaan door de eenzaamheid van de mensen, maar ook door de wanhoop van het leiden van een ongezien leven. De gedachte dat niemand wist wanneer ze thuiskwamen, naar bed gingen en wakker werden, teisterde hen. Veel mensen houden een sterk onbevredigende relatie gaande omdat ze snakken naar een getuige van hun leven als buffer tegen existentiële isolatie. Onze behoefte naar verbinding en contact en ons besef ten diepste alleen te zijn in de wereld zijn in conflict met elkaar. Een diep gevoel van verbinding lost niet zomaar de existentiële isolatie op, maar het biedt wel troost.³⁵ Mensen zijn sociale dieren. Chronische eenzaamheid is zelden een keuze.

De angst voor existentiële isolatie wordt waarneembaar als mensen erbij stilstaan dat er momenten zijn waarop er niemand op de wereld aan hen denkt. Dat zijn de momenten waarop we snakken naar Facebooknotificaties, apps, softwarebots, hologrammen, slimme thermostaten en slimme koelkasten die beginnen te praten. Maar ook de radio op de achtergrond of tv-kijken kan zo'n gevoel van isolatie verminderen. Je zou kunnen zeggen dat met de smartphone in de hand een einde is

35 De existentiële angst voor sociale isolatie slaat op het besef dat we uiteindelijk alleen geboren worden en alleen dood gaan. Dit besef gaat niet weg door sociale verbinding, maar wordt er wel door verzacht. Zie: Wedding, D. & J.C. Raymond, *Current Psychotherapies*, Brooks/Cole, 2014.

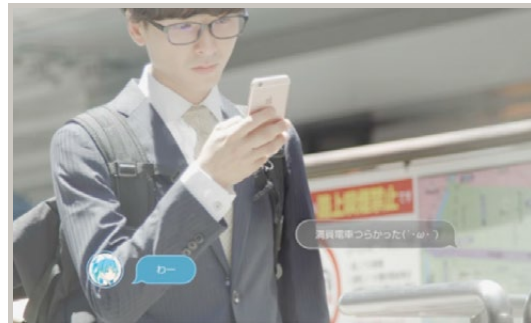
'I mean, they say you die twice. One time when you stop breathing and a second time, a bit later on, when somebody says your name for the last time.'

Graffiti-artiest Banksy³⁶

gekomen aan een ongezien leven. Google rapporteert dat er vorig jaar 24 miljard selfies zijn geüpload naar hun foto-app.³⁷ Iedere seconde worden er 9000 nieuwe foto's geüpload op Snapchat (een kleine 300 miljard per jaar), waarvan de meeste selfies zijn. Instagram, Facebook, Baidu en andere media moeten daar nog even bij worden opgeteld.

Het Japanse bedrijf Vinclu geeft een ontluisterend beeld van de nieuwe mogelijkheden. Hun AI-product Gatebox is te koop voor 2500 euro en bestaat uit een vrouwelijk hologram, qua functie en features een chatbot die je ook in een product als Amazon Echo terugvindt. In het promotiefilmpje zien we dat het hologram chat met een jongeman die naar zijn werk gaat. Tijdens de lunch krijgt hij het bericht van de chatbot dat het hologram hem mist en of hij niet wat eerder thuis kan komen die avond. Eenmaal thuis verzucht de man: 'Somebody's home for me. Feels great.'³⁸

Het is duidelijk een surrogaat, maar machinegevoelens kunnen ook echt beleefd worden. De gevoelens die militairen



Voor 2.500 euro verdrijft het chattende hologram van Vinclu jouw eenzaamheid.

in oorlogssituaties ontwikkelen voor hun robothulpjes die ze inzetten in gevaarlijke situaties, is hier een goed voorbeeld van. Mensen bouwen een band op met die machines, zegt defensiespecialist Peter Singer.³⁹ Er zijn gevallen bekend waarin de militair zijn leven waagt om de robot te redden, de robot een militaire promotie

36 <http://legacymultimedia.com/2015/01/09/when-is-your-memory-truly-forgotten/>

37 <http://www.dailymail.co.uk/sciencetech/article-3619679/What-vain-bunch-really-24-billion-selfies-uploaded-Google-year.html>

38 <https://www.youtube.com/watch?v=nkcKaNgfykg&feature=youtu.be>

39 Singer, P., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin Press HC, 2009.

'If they can give the appearance of aliveness and yet not disappoint, they may even have a comparative advantage over people, and open new possibilities for narcissistic experience with machines.'

Sherry Turkle

krijgt of soldaten een begrafenis organiseren als het materieel niet meer te repareren is. We refereren hier weer even naar ons vorige rapport over machine-intelligentie (*Het bot-effect: friending your brand*), waarin we de antropomorfe relatie en de emotionele connectie met AI beschreven. In deze surrogaatoplossing schuilt natuurlijk een dilemma.

Sherry Turkle, professor in Social Studies of Science and Technology aan MIT, waarschuwt dat virtuele placebo's leiden tot de vervlakking van intermenselijk contact. In haar boek *Alone Together: Why We Expect More from Technology and Less from Each Other*,⁴⁰ gaat het met name nog om het gebruik van de smartphone. Die zorgt ervoor dat we ons isoleren van anderen doordat we meer aandacht geven aan de smartphone dan aan de mensen om ons heen. Tegelijkertijd zijn er vele studies die bevestigen dat de eenzaamheid van ouderen door de inzet van zorg-

robots minder wordt.⁴¹ Wat dat betreft een goede zaak dus, die robots. In een studie die Turkle uitvoerde naar zorgrobots⁴² concludeert ze als keerzijde dat de inzet van robots kan leiden tot meer narcisme, mensen die meer met zichzelf bezig zijn dan met anderen, maar ook dat robots een competitief voordeel hebben.

Onderzoekers van de Universiteit van Sheffield wijzen in hun studie 'Granny and the robots'⁴³ ook op de potentiële gevaren, zoals:

- mogelijke reductie van de hoeveelheid menselijk contact;
- toename van de kans op objectificatie van de patiënt: de patiënt als 'ding' gaan zien;
- verlies van privacy (bijvoorbeeld omdat de robot data opslaat of anderen mee kunnen kijken);
- verlies van persoonlijke vrijheid;
- deceptie en infantilisering.

40 Turkle, S., *Alone Together: Why We Expect More from Technology and Less from Each Other*, Basic Books, 2011.

41 Zie bijvoorbeeld: Robinson, H. e.a., *The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial*, 2013.

42 Taggart, W., S. Turkle & C. Kidd, 'An Interactive Robot in a Nursing Home: Preliminary Remarks', 2005, <http://www.androidscience.com/proceedings2005/TaggartCogSci2005AS.pdf>

43 Sharkey, A. & N. Sharkey, 'Granny and the robots: ethical issues in robot care for elderly', 2010, <http://link.springer.com/article/10.1007/s10676-010-9234-6>

Als robots ons gaan appen en zeggen dat zij zich alleen voelen en blij zijn als je weer snel thuiskomt, dan is dat natuurlijk de wereld op zijn kop. Natuurlijk kun je ook beargumenteren dat AI juist ervoor zorgt dat er meer connecties met mensen worden gelegd. Facebook attendeert je op de verjaardagen van je vrienden, LinkedIn doet suggesties om met mensen die je misschien kent een connectie te maken. Of wat te denken van Bernie, de AI dating app die het swipe-werk en de eerste openingszinnen met potentiële partners voor zijn rekening neemt. Allemaal op basis van het leren begrijpen van je voorkeuren en smaak. Ideaal voor wie de casanova wil uithangen, maar geen tijd heeft om iedereen af te struinen.

Tot slot: het doorbreken van de beschermende cocon

De Big Four zijn herkenbaar maar ook benauwend; we willen meestal helemaal niet aan deze angsten denken. Socioloog Anthony Giddens vertelt in zijn boek *Modernity and Self-Identity* hoe we een veilige cocon om ons heen spinnen waarmee we onze angsten en bedreigingen buitensluiten. We ontwikkelen mechanismen en denkstrategieën om onszelf veilig te wanen en vertrouwen erop dat de wereld 'normaal' en voorspelbaar is. Dit kunnen we vaak lange tijd volhouden, maar soms is een gebeurtenis zo doordringend dat de cocon breekt en we oog in oog staan met onze angst.

Door haar menselijke eigenschappen en vervlechting met onze levens zorgt AI voor extra confrontatiemomenten. Wanneer we een kunstmatige intelligentie creëren die het oneindige leven heeft, moeten we onze eigen sterfelijkheid onder ogen zien. Onze veilige cocon wordt op zo'n moment doorbroken; we worden geconfronteerd met onze angst om zelf dood te gaan.

Wanneer we zien hoe de mensen om ons heen steeds meer tijd aan hun mobiel of hun zorgrobot besteden, worden we geconfronteerd met onze angst om buitengesloten te worden. Op deze manier doorbreekt AI keer op keer onze beschermende cocon. AI werkt hierbij ook als spiegel: het toont ons onze diepste angsten en stelt ons voor het blok. Belangrijke levensvragen dringen zich ongevraagd op: wat betekent de dood? Willen we een oneindig leven? Wat betekent 'echt contact'? Kunnen we eigenlijk over alles controle hebben? En wat is nou de betekenis van mijn leven?

Yalom juicht dergelijke confrontaties toe en ziet dit als enige manier om constructief met de angsten om te gaan. Niet met als doel om nooit meer bang te zijn, maar in ieder geval om een betere verhouding met onze angsten te vinden. Reden dus om de angsten niet uit de weg te gaan en het gesprek met uw werknemers en klanten aan te gaan.

DE FRANKENSTEINFACTOR: CIO ALS THERAPEUT

'In any psychotherapy, the therapist himself is a highly important part of the human equation. What he does, the attitude he holds, his basic concept of his role, all influence therapy to a marked degree.'⁴⁴

Carl Rogers, in zijn boek *Client Centred Therapy*

De angst voor superslimheid en superdomheid, de angst van ons onderbewuste, de angst voor de achteruitgang van cultuur en natuur en de vier levensangsten waar mensen altijd mee worstelen zijn vier perspectieven om de angst te herkennen,

te erkennen en te analyseren. We zetten hier voor het gemak staccato factor Frankenstein 1 tot en met Frankenstein 4 nogmaals op een rijtje en leggen de onderlinge verbanden.

Frankenstein 1 Angst voor slimheid en domheid

Deze twee angstuitersten hebben dezelfde oorsprong. De AI-toepassingen krijgen meer autonomie en gaan zelfstandig handelen. Op korte termijn gaat het mis vanwege gebrek aan intelligentie. Op lange termijn gaat het mis omdat ze te intelligent zijn. De onderliggende angst is hetzelfde: we hebben geen controle over de AI-machinerie en Frankenstein keert zich tegen ons.

Frankenstein 2 Angst voor de 'uncanny'

Het gevoel, met name de onderbuik, regeert bij de confrontatie met de onbekende ander, de automaton of androïde. De Frankenstein die we hebben gebouwd, zijn we zelf, en we kennen onszelf niet erg goed. Het kwaad dat in ieder mens schuilt projecteren we op de digitale dubbelganger.

⁴⁴ Chapter 2, https://books.google.nl/books?hl=nl&lr=&id=bWieBAAQBAJ&oi=fnd&pg=PT13&dq=therapy+roleplaying&ots=LwitLxFLi8&sig=gC_tFNqnZmcnal-YAU-YjqOoqs8#v=onepage&q=Play%20therapy&f=false

Frankenstein 3 Angst voor achteruitgang van cultuur en natuur

Reserveringen ten aanzien van technologische vooruitgang zitten geworteld in de westerse en oosterse cultuur. Beide worstelen met de vraag hoe ver technologie mag oprukken en het leven mag veranderen. Het temmen van de technologie en het overwinnen van de natuur tonen de karakters van Oost en West. Na het overwinnen van de natuur kan die natuur zich tegen ons keren. Dat verhaal ontstaat uit de confrontatie van de verlichting en de romantiek, die in strijd met elkaar zijn. Het is de moraal van het Frankensteinverhaal.

Frankenstein 4 Angst voor de Big Four

Als we met de huidige stand van de AI-techniek nieuwe Frankensteins gaan bouwen, welke eigenschappen geven we ze dan mee? Producten die mensen in het zonnetje zetten, ze erkennen en gezien laten worden, producten die mensen verbinden en uit hun isolement halen, die betekenis geven aan het leven en ons de controle in handen geven, of andersom? Bij de andersomversie staren we de dood bijna letterlijk in de ogen en hebben we feitelijk een echte Frankenstein gebouwd.

'The task of the therapist is to reduce anxiety to comfortable levels and then to use this existing anxiety to increase a patient's awareness and vitality.'⁴⁵

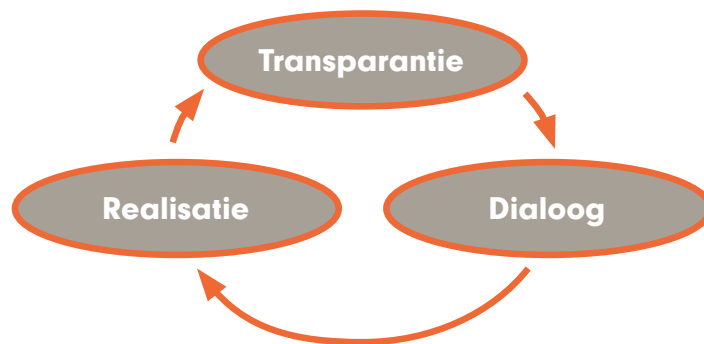
Irvin Yalom

De CIO staat niet alleen in de uitdaging om het maximale uit AI te halen. Zijn stakeholders, de klant, de werknemers, zijn peers, de marketeers en de aandeelhouders hebben immers allemaal baat bij de vruchtbare inzet van kunstmatige intelligentie. Het gaat dus om de gehele organisatie en het merk, en de vraag is welke rol de CIO moet spelen. Is hij de volger, de puur technologisch gedreven enabler, of de initiator en degene die het voortouw neemt? Sommigen zeggen dat het logischer is dat de CTO, Chief Digital Officer of Chief Marketing Officer zich hierom bekommert. Het zal per organisatie en mate van volwassenheid verschillen, maar het zal duidelijk zijn dat wij met dit rapport de CIO een duwtje in de rug willen geven.

45 Yalom, I.D., *Existential Therapy*, Basic Books, 1980, p. 188.

Maar wie het ook doet, het is een debat dat nu gevoerd moet worden. In het ergste geval zoekt de klant zijn business ergens anders, waar zijn angsten wel serieus worden genomen en vertrekt de werknemer naar een concurrent die beter omgaat met de emotie en het onderbuikgevoel van zijn werknemers. Wat staat de CIO dan te doen? Een eenvoudige ABC-formule, zoals gesuggereerd aan het begin van dit

rapport, staat per definitie op gespannen voet met de heftige materie die we hier beschrijven. Tegelijkertijd is het duidelijk dat er angst heerst en dat deze geadresseerd moet worden. De metafoor 'de CIO als therapeut' biedt uitkomst: in therapie zijn ook geen eenvoudige hapklare oplossingen, wel voorwaarden voor een vruchtbare sessie:



A. Transparantie. Het wordt genoemd in de rapporten van de EU, door Luciano Floridi en ook in de 23 AI-principes van Elon Musk en Stephen Hawking komt transparantie terug. AI is omgeven door speculatie en transparantie kan gezien worden als 'het licht aandoen in een donkere kamer'. Voor de CIO betekent dit transparant zijn in plannen, activiteiten en intenties die de organisatie heeft met AI. Maar ook om alvast in lijn te opereren met de Europese resolutie die in februari 2017 is aangenomen, waarin transparantie een centrale rol speelt. Denk ook aan het bespreekbaar maken van de werkgelegenheidsgevolgen en de noodzaak tot omscholing. Waar moeten werknemers nou wel en niet rekening mee houden? Voor de klant kan het helpen wanneer duidelijk wordt vermeld wat de AI wel en niet kan; ook hier krijgen angsten en fantasieën zo minder speelruimte. We noemden al Mercedes-Benz dat kenbaar maakte voor wie hun zelfrijdende auto zou kiezen wanneer een ongeluk onvermijdelijk is. Een transparantieknop kan ook helpen: wanneer hierop gedrukt wordt, wordt de argumentatie van de AI uitgelegd en wordt duidelijk waarom er een bepaalde keuze is gemaakt.

B. Dialoog. De transparantie geeft stof voor, en opent de weg naar, de dialoog. Dat mag volgens de nieuwe standaarden in de psychologie ook door het tonen van de eigen gevoelens en onzekerheden. Yalom verwijst daarmee het klassieke beeld waarin de therapeut luistert en de patiënt op de sofa praat tot het verleden. Het biedt de mogelijkheid 'echt' in contact te komen met de mens achter het product en het systeem. Naast efficiëntie en effectiviteit dus ook de existentie in de dialoog betrekken: wat doet AI met u, uw werknemers en uw klanten? Uw angstkompas, dat hopelijk is aangescherpt door dit rapport, helpt u de angsten te herkennen en te benoemen. Een diepgaand begrip van uw werknemer en klant: het kan uw voorsprong op de concurrent zijn. In de huid kruipen van de stakeholders en een analyse maken van de vier F-factoren vanuit hun perspectief is een goede eerste stap.

Experimenteer in de voortgaande dialoog met uw klant. 'Exposure in vivo' is bijvoorbeeld een psychologische methode waarbij stap voor stap gewend wordt aan situaties die eng gevonden worden. De methode wordt onder andere gebruikt tegen angstfobieën.⁴⁶ Voor doelgroepen waarbij bijvoorbeeld de angst voor controleverlies opspeelt omdat het tempo van technologische ontwikkelingen te hoog ligt, kan dit de oplossing zijn: bedenk 'tussenfases' of tussenproducten waarbij niet alle controle weggenomen wordt. De 'Stewart' als mediator tussen de zelfrijdende auto en de inzittende is daarvan een goed voorbeeld.

C. Realisatie. Het toepassen van de inzichten uit deze existentiële dialoog in het beleid en in uw producten. De inzichten uit de analyse van de FrankensteinFactor wijzen al in een bepaalde richting. Met inachtneming van de Big Four bent u in staat de meest betrouwbare, waardevolle en geliefde IT-producten en -diensten te maken die u zich maar kunt wensen. Geen Frankenstein die ons doet vreemden of vervlakken, maar nieuwe mens-machinerelaties die het leven meer betekenis geven en en passant uw organisatie purpose-driven maken. Laat AI uw kernwaarden ondersteunen en de menselijkheid niet ondermijnen maar juist versterken.

Het zijn abstracte handvatten en het echte werk zult u, net als altijd, zelf moeten doen. Houd in ieder geval uw angstkompas bij de hand en adresseer een angst wanneer

u die herkent. Doe als CIO de lamp aan in de donkere kamer, dan valt het misschien nog wel mee met het monster onder het bed.

NAWOORD: ANGST IS DE ZEITGEIST

Er is nog een argument voor de anatomie van de angst dat we nog hebben laten liggen. Als angst in zijn algemeenheid een trend wordt, en er dus makkelijker een angsetiket op zaken wordt geplakt, dan is angst voor AI niet anders dan een reflectie van de huidige zeitgeist. Dominique Moïsi, die oprichter is van de vermaarde Franse denktank Institut Français des Relations Internationales (IFRI), stelt dat de angst tegenwoordig regeert. Hij maakte een nieuwe angstanalyse na het bekijken van populaire series op Netflix, een project dat hij 'du bingewatching' noemt. Die tv-series zijn volgens Moïsi een weergave van het huidige sentiment en weerspiegelen onze angsten. Hij signaleert er vijf die in de huidige maatschappij de boventoon voeren.

- angst voor chaos en barbarij (*Game of Thrones*);
- angst voor verval van de democratie (*House of Cards*);
- angst voor terreur of het niet kennen van de identiteit van de vijand (*Homeland*);
- angst voor het verdwijnen van de bestaande orde (*Downton Abbey*);
- angst voor een Russische bezetting (*Occupied*).

Twee populaire robotseries zouden wij graag aan deze lijst toevoegen:

- angst voor dominantie van robots over de mens (*Real Humans*);
- angst voor dominantie van de mens over robots en angst voor onze eigen slechtheid (*Westworld*).

Als het inderdaad zo is dat de angst in deze populaire tv-series een spiegel vormt van de maatschappij, dan is de angst voor AI slechts onderdeel van een groter verhaal. De aanwezige angst wordt volgens Moïsi gevoed door het centrale thema voor deze 21e eeuw: wie zijn we, wie willen we zijn en hoe bevechten we die identiteit? De robotseriethematiek past daar naadloos in. Nu AI alles voor ons kan doen en kan gaan betekenen, wat betekenen wij dan nog en hoe bepalen wij onze positie ten opzichte van onze dubbelgangers?

BIJLAGE A

ASILOMAR AI PRINCIPLES (2017)

Onderzoekskwesties

1. Onderzoeksdoel: het doel van AI-onderzoek zou moeten zijn om niet ongerichte intelligentie te creëren, maar intelligentie die een gunstig effect heeft.
2. Onderzoeksfinanciering: investeringen in AI moeten gepaard gaan met de financiering van onderzoek dat het gunstige gebruik waarborgt, met inbegrip van netelige vragen in de informatica, economie, recht, ethiek en sociale studies, zoals:
 - Hoe kunnen we toekomstige AI-systemen zeer robuust maken, zodat ze doen wat we willen zonder een storing te krijgen of gehackt te worden?
 - Hoe kunnen we onze welvaart laten groeien door middel van automatisering met behoud van de middelen en het doel van mensen?
 - Hoe kunnen we onze rechtssystemen eerlijk en efficiënter maken en gelijke tred laten houden met AI en de risico's in verband met AI beheren?
 - Op welke set van waarden moet AI worden afgestemd en welke juridische en ethische status moet AI hebben?
3. Link tussen wetenschap en politiek: er moet een constructieve en gezonde uitwisseling plaatsvinden tussen AI-onderzoekers en beleidsmakers.
4. Onderzoekscultuur: een cultuur van samenwerking, vertrouwen en transparantie dient te worden bevorderd tussen onderzoekers en ontwikkelaars van AI.
5. Snelheidsbeperking: teams die AI-systemen ontwikkelen, moeten actief samenwerken om te voorkomen dat het niet zo breed wordt genomen met de veiligheidsnormen.

Ethiek en waarden

6. Veiligheid: AI-systemen moeten veilig zijn tijdens de hele operationele levensduur en controleerbaar waar van toepassing en haalbaar.
7. Fouttransparantie: als een AI-systeem schade berokkent, moet het mogelijk zijn om na te gaan waarom.
8. Juridische transparantie: elke betrokkenheid van een autonoom systeem in gerechtelijke besluitvorming zou een bevredigende verklaring moeten geven die controleerbaar is door een bevoegde menselijke autoriteit.
9. Verantwoordelijkheid: ontwerpers en bouwers van geavanceerde AI-systemen zijn stakeholders in de morele implicaties van gebruik, misbruik en acties, met een verantwoordelijkheid en kans om die gevolgen vorm te geven.
10. Waarde-afstemming: hoog autonome AI-systemen moeten zo ontworpen zijn dat kan worden verzekerd dat hun doelen en gedragingen tijdens hun werking zijn afgestemd op menselijke waarden.

11. Menselijke waarden: AI-systemen moeten zijn ontworpen en moeten functioneren om in overeenstemming te zijn met idealen van de menselijke waardigheid, rechten, vrijheden en culturele diversiteit.
12. Persoonlijke levenssfeer: mensen moeten het recht hebben op toegang tot en op het managen en beheren van de gegevens die ze genereren, gegeven de mogelijkheden van AI-systemen om deze data te analyseren en te gebruiken.
13. Vrijheid en privacy: het toepassen van AI op persoonsgegevens mag niet op onredelijke wijze de echte of vermeende vrijheid van mensen inperken.
14. Gezamenlijk voordeel: AI-technologieën zouden de positie van zo veel mogelijk mensen ten goede moeten komen en moeten versterken.
15. Gedeelde welvaart: de economische welvaart gecreëerd door de AI moet breed worden gedeeld, om de hele mensheid ten goede komen.
16. Menselijke controle: mensen moeten kiezen hoe en of beslissingen worden gedelegeerd aan AI-systemen om de door mensen gekozen doelstellingen te realiseren.
17. Niet-ondermijning: de macht die aan zeer geavanceerde AI-systemen is toegekend, moet de sociale en maatschappelijke processen die de gezondheid van de samenleving bepalen, respecteren en verbeteren in plaats van ondermijnen.
18. AI-wapenwedloop: een wapenwedloop van dodelijke autonome wapens moet worden vermeden.

Langetermijnkwesties

19. Capabiliteitsvoorzichtigheid: aangezien er geen consensus is, moeten we sterke aannames ten aanzien van bovengrenzen voor toekomstige AI-mogelijkheden voorkomen.
20. Belang: geavanceerde AI zou een ingrijpende verandering in de geschiedenis van het leven op aarde teweeg kunnen brengen en zou worden gepland en beheerd met evenredige zorg en middelen.
21. Risico's: risico's van AI-systemen, met name catastrofische of existentiële risico's, moeten worden onderworpen aan planning en reductie-inspanningen in overeenstemming met hun verwachte impact.
22. Recursieve zelfverbetering: AI-systemen die zijn ontworpen om recursief zichzelf te verbeteren of te repliceren, op een wijze die zou kunnen leiden tot snel toenemende kwaliteit of kwantiteit, moeten worden onderworpen aan strenge veiligheids- en controlemaatregelen.
23. Algemeen belang: superintelligentie mag alleen worden ontwikkeld in dienst van de breed gedeelde ethische idealen, en ten behoeve van de hele mensheid in plaats van een staat of organisatie.

BIJLAGE B

FRANKENSTEIN SYNDROME

QUESTIONNAIRE

1. Wijdverspreid gebruik van menselijke robots zal banen van mensen wegnemen.
2. Menselijke robots maken ons luier.
3. Als een robot schade aan mensen toebrengt, zouden de personen en organisaties die die robots hebben gemaakt, de slachtoffers voldoende moeten compenseren.
4. Ik ben bang dat robots minder contact tussen mensen zullen aanmoedigen.
5. Ik ben bang dat er iets kan gebeuren als we te afhankelijk worden van robots.
6. Ik haat het idee dat robots of kunstmatige intelligentie over bepaalde zaken zouden gaan oordelen.
7. Ik zou me ongemakkelijk voelen als robots echt emoties hadden en eigen gedachten konden vormen.
8. Er zou iets slechts kunnen gebeuren als menselijke robots zich zouden ontwikkelen tot mensen.
9. Menselijke robots zouden gevaarlijke taken moeten uitvoeren, zoals werken in rampgebieden, diep in zee of buiten deze planeet.
10. Ik ben bang dat robots een slechte invloed zullen hebben op kinderen.
11. Heel veel menselijke robots in de maatschappij zullen het minder warm maken.
12. Ik kan een persoon of organisatie die robots maakt vertrouwen.
13. Menselijke robots kunnen heel nuttig zijn om onze kinderen wat te leren.
14. Ik weet niet waarom, maar ik vind het idee van menselijke robots wel leuk.
15. Ik vertrouw mensen en organisaties die robots bouwen dat ze voldoende informatie met de buitenwereld zullen delen, inclusief negatieve informatie.
16. Menselijke robots kunnen ons leven makkelijker maken.
17. Personen en organisaties die robots maken, bedoelen het goed.
18. Menselijke robots zijn een natuurlijk product van onze beschaving.
19. Personen en organisaties die de robots maken, houden rekening met de behoeften, gedachten en gevoelens van de gebruikers.
20. Menselijke robots zijn erg nuttig voor oudere mensen en gehandicapten.
21. Menselijke robots zouden repetitieve saaie taken uit moeten voeren in plaats van deze aan mensen over te laten.
22. De ontwikkeling van menselijke robots is godslastering.
23. De ontwikkeling van menselijke robots is blasfemie tegen de natuur.
24. De technologieën die nodig zijn voor de ontwikkeling van menselijke robots, horen bij een wetenschappelijk veld dat we niet zouden moeten bestuderen.

25. Ik weet niet waarom, maar robots maken me bang.
26. Ik voel dat de maatschappij in de toekomst gedomineerd gaat worden door robots.
27. Ik ben bang dat menselijke robots ons doen vergeten wat het is om mens te zijn.
28. Als mensen met menselijke robots interacteren, zou dat soms kunnen leiden tot problemen in de relatie tussen mensen.
29. Menselijke robots kunnen nieuwe vormen van communicatie creëren, zowel tussen mensen als tussen mensen en machines.
30. Een wijdverspreid gebruik van robots zou betekenen dat het duur wordt om ze te behouden.

- Adams, T., 'Artificial intelligence: "We're like children playing with a bomb"', *The Guardian*, 12 juni 2016
<https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>
- Arntz, M., T. Gregory & U. Zierahn, 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis', OECD Social, Employment and Migration Working Papers, 14 mei 2016
http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en
- Amodei, D., e.a., 'Concrete Problems in AI Safety', Cornell University Library, 26 juli 2016 <https://arxiv.org/abs/1606.06565>
- Bartneck, C., 'Who like androids more: Japanese or US Americans', Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, 1-3 augustus 2008
<https://www.bartneck.de/publications/2008/whoLikesAndroids/bartneckRoMan2008.pdf>
- Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 3 juli 2014
- British Science Association, 'One in three believe that the rise of artificial intelligence is a threat to humanity'
<http://www.britishtscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity>
- Chapman University, 'America's Top Fears 2015', 13 oktober 2015
<https://blogs.chapman.edu/wilkinson/2015/10/13/americas-top-fears-2015/>
- Deloitte, 'Technology and people: The great job-creating machine', 2015
<https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf>
- Doland, A., 'Check Out the First Ad From McCann Japan's "AI Creative Director"', *AdvertisingAge*, 6 juni 2016
<http://adage.com/article/creativity/check-commercial-mccann-japan-ai-creative-director/304320/>
- Emmelkamp, P.M.G. e.a., 'Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia', *Behaviour Research and Therapy* 40(5), mei 2002
<http://www.sciencedirect.com/science/article/pii/S0005796701000237>
- European Parliament, Committee on Legal Affairs, *DRAFT REPORT with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, 31 mei 2016 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN>

- Floridi, L., 'Should we be afraid of AI?', *Aeon*, 9 mei 2016
<https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>
- Forrester Research, 'Understanding the role of emotion in customer experience', 2015 <https://www.forrester.com/report/Understanding+The+Impact+Of+Emotion+On+Customer+Experience/-/E-RES122503>
- Frankl, V., *Man's Search for Meaning*, Beacon Press, 1 juni 2006 (eerste publicatie: 1946)
- Frey, C. & A. Osborne, 'The Future of Deployment', 2013
http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- Future of Life Institute, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', 2015
<http://futureoflife.org/open-letter-autonomous-weapons/>
- Godwin, W., *Lives of the Necromancers or The Exercise of Magical Power*, 1876, p. 180
<https://archive.org/details/livesnecromance04godwgoog>
- Gooch, J.J., 'NVIDIA Joins Cancer Moonshot with AI Platform for Accelerating Research', *Digital Engineering*, 14 november 2016
<http://www.digitaleng.news/de/nvidia-joins-cancer-moonshot-with-ai-platform-for-accelerating-research/>
- Gray, R., 'What a vain bunch we really are! 24 billion selfies were uploaded to Google last year', *MailOnline*, 1 juni 2016
<http://www.dailymail.co.uk/sciencetech/article-3619679/What-vain-bunch-really-24-billion-selfies-uploaded-Google-year.html>
- Hayasaki, E., 'Is AI Sexist?' *Foreign Policy*, januari 2017
<http://foreignpolicy.com/2017/01/16/women-vs-the-machine/>
- House of Commons Science and Technology Committee, *Robotics and artificial intelligence*, september 2016
<https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>
- Lewis, D.A., G. Blum & N.K. Modirzadeh, 'War-Algorithm Accountability', 31 augustus 2016
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2832734
- McKinsey Global Institute, 'Harnessing automation for a future that works', januari 2017
<http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>
- Misener, D. & D. Maloney, 'Wired founder Kevin Kelly on letting go of AI anxiety', 6 december 2016
<https://slackhq.com/wired-founder-kevin-kelly-on-letting-go-of-ai-anxiety-bcd94e50dedc#amf4qbyxp>
- Nomura¹, T., D.S. Syrdal & K. Dautenhahn, 'Differences on Social Acceptance of Humanoid Robots between Japan and the UK', 4th International Symposium on New Frontiers in Human-Robot Interaction, 21-22 april 2015
http://uhra.herts.ac.uk/bitstream/handle/2299/16345/Nomura_etal_AISB2015_differencesonsocial.pdf;sequence=2

- Robinson, H. e.a., *The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial*, 2013
- Rogers, C., *Client Centred Therapy*, 1951
- Ros, F., Stewart II, juni 2016
http://felixros.com/projects.html#big_wrap
- Sharkey, A. & N. Sharkey, 'Granny and the robots: ethical issues in robot care for elderly', 2010
<http://link.springer.com/article/10.1007/s10676-010-9234-6>
- Shaughnessy, M.F., D. Main, J. Madewell & I. Yalom, 'An interview with Irvin Yalom', *North American Journal of Psychology* 9(3), december 2007 <http://www.freepatentsonline.com/article/North-American-Journal-Psychology/173513562.html>
- Singer, P., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin Press HC, 2009
- Stewart, J., 'Strong Artificial Intelligence and National Security: Operational and Strategic Implications', U.S. Joint Military Operations Department, 18 mei 2015
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA622591>
- Taggart, W., S. Turkle & C. Kidd, 'An Interactive Robot in a Nursing Home: Preliminary Remarks', 2005
<http://www.androidscience.com/proceedings2005/TaggartCogSci2005AS.pdf>
- Turkle, S., *Alone Together: Why We Expect More from Technology and Less from Each Other*, Basic Books, 2011
- Twyford, S., 'When Is Your Memory Truly Forgotten?', *Legacy Multimedia*, 9 januari 2015
<http://legacymultimedia.com/2015/01/09/when-is-your-memory-truly-forgotten/>
- Vinclu, 'Gatebox - Virtual Home Robot', YouTube, 13 december 2016
<https://www.youtube.com/watch?v=nkcKaNqfykg&feature=youtu.be>
- Walsh, J. & M. Schulzke, 'The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?', United States Army War College, september 2015
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA621793>
- Wedding, D. & J.C. Raymond, *Current Psychotherapies*, Brooks/Cole, 2014
- Woollaston, V., 'Do YOU trust driverless cars? Only 1 in 10 of us feel confident about getting in autonomous vehicles and women are the most hesitant', *MailOnline*, 23 september 2015
<http://www.dailymail.co.uk/sciencetech/article-3245058/Do-trust-driverless-cars-1-10-feel-confident-getting-autonomous-vehicles-women-skeptical.html>
- World Economic Forum, *The Global Risks Report 2017*
<http://reports.weforum.org/global-risks-2017/executive-summary/>
- Yalom, I.D., *Existential Psychotherapy*, Basic Books, 1980

Over VINT labs.sogeti.com

Het Verkenninginstituut Nieuwe Technologie van Sogeti, VINT, is onderdeel van SogetiLabs. VINT geeft invulling aan de koppeling tussen bedrijfsprocessen en nieuwe IT. In elke rapportage over een verkenning die het instituut heeft uitgevoerd, zoekt VINT het juiste midden tussen feitelijke beschrijving en beoogde toepassing. Op die manier inspireert VINT organisaties om nieuwe technologie in beschouwing te nemen of zelfs te gaan gebruiken.

De onderzoeken van VINT worden uitgevoerd onder auspiciën van de Commissie van Aanbeveling bestaande uit:

- F.M.R. van der Horst (voorzitter), CEO Retail Banking ABN AMRO Bank N.V.
- J. Behrens, Senior VP, Business Leader Personal Health Solutions Philips
- P. Dirix, Managing Director Operations NS Reizigers
- D. Kamst, Chief Executive Officer Klooker
- Prof. dr. ir. R. Maes, Professor of Information & Communication Management Academy for I & M
- P. Morley, Member Supervisory Board TLS
- T. van der Linden, Group Information Officer Achmea
- E. Schuchmann, Chief Information Officer Academisch Medisch Centrum
- K. Smaling, Chief Technology Officer Continental Europe Aegon
- W.H.G. Sijstermans, Member of the Board, CIO Nederlandse Zorgautoriteit (NZa)
- M. der Weduwen, CIO & Vice President Information Services, Holland Casino
- M. Boreel, Chief Technology Officer Sogeti Group
- J.P.E. van Waayenburg, Chief Executive Officer Sogeti Group
- P.W. Wagter, Chief Executive Officer Sogeti Nederland B.V.

Over SogetiLabs labs.sogeti.com

SogetiLabs is een netwerk van meer dan 120 technologie-leiders binnen Sogeti wereldwijd. SogetiLabs biedt een breed scala aan deskundigheid op het gebied van digitale technologie: van embedded software, cyber security, simulaties en cloud tot business information management, mobiele apps, business analytics, testen en het Internet of Things. Het doel is altijd om technologie, systemen en applicaties met een zo groot mogelijk resultaat in actuele bedrijfssituaties in te zetten. Samen met VINT zorgt SogetiLabs voor inzicht, onderzoek en inspiratie met artikelen, presentaties en video's. Deze zijn altijd te downloaden op de uitgebreide website, online portals en social media van SogetiLabs.

Over Sogeti Nederland B.V.

Sogeti is een toonaangevende IT-dienstverlener op het gebied van technologie en engineering services. Het bedrijf biedt oplossingen die digitale transformatie mogelijk maken. Grootzakelijke organisaties vertrouwen op de uitgebreide kennis en ervaring voor clouddienstverlening, digitale veiligheid, kwaliteitszorg & testing en digital manufacturing. Sogeti combineert daarbij slagvaardigheid en implementatiesnelheid met sterke technologiepartnerships volgens bewezen methodieken. De IT-dienstverlener voorziet in een eigen wereldwijd leveringsmodel Rightshore®.

Passie voor het vakmanschap zit in de genen van de meer dan 2450 Sogetisten werkzaam (dicht)bij de klant. Sogeti Nederland B.V. maakt deel uit van de wereldwijde Sogeti groep met zo'n 25.000 IT-professionals in 15 landen, gevestigd in meer dan 100 locaties in Europa, de VS en India. De Sogeti groep is een dochteronderneming van het in Parijs beursgenoteerde bedrijf Cap Gemini S.A. Meer informatie is beschikbaar op www.sogeti.nl.

