



Data is macht

Over Big Data en de toekomst

Dhoya Snijders, redactie



Colofon

Redactie en Projectleiding: Dhoya Snijders

Tekst- en taalredactie: Annette Potting, Japke Schreuders

Met medewerking van: diverse auteurs

Vormgeving en illustraties: Ellen Bouma grafisch vormgeven en dtp

Drukwerk: Quantes

Afbeeldingen: De kunstwerken in deze publicatie en op de kaft zijn gemaakt door het zelflerend neurale netwerk *Inceptionism* van Google Research.

ISBN 978-94-91397-15-8

STT-publicatie nr. 86

NUR 950

Trefwoorden: Big Data, digitalisering, kunstmatige intelligentie, toekomstvisie, technologische ontwikkelingen, maatschappij.

© 2017, Stichting Toekomstbeeld der Techniek, Den Haag



Data is Macht. Over Big Data en de toekomst van Stichting Toekomstbeeld der Techniek wordt auteursrechtelijk beschermd zoals vastgelegd onder de Creative Commons Naamsvermelding Niet Commercieel-Geen Afgeleide Werken 3.0 Unported licentie. U kunt dit werk toeschrijven aan Stichting Toekomstbeeld der Techniek / Dhoya Snijders (www.stt.nl), 2017. Bezoek www.creativecommons.org/licenses/by-nc-nd/3.0 voor de volledige tekst van de licentie

Stichting Toekomstbeeld der Techniek

Prinsessegracht 23, 2514 AP Den Haag

Postbus 30424, 2500 GK Den Haag

070-302 98 30

info@stt.nl

www.stt.nl

Data is macht

Over Big Data en de toekomst

Stichting Toekomstbeeld der Techniek

Dhoya Snijders, redactie

INHOUDSOPGAVE

Voorwoord	6
Samenvatting	8
1. Data als DNA	13
2. De staat van data	31
3. Wat telt? Over data en sociale gevolgen	53
4. Tijdsdimensies: willen we meer of minder data?	79
5. Nieuwe verhalen	99
6. Wat nu?	119
Bijlagen	123
Aanpak	123
Dankwoord en betrokkenen	124
Stichting Toekomstbeeld der Techniek	125
Literatuur	128
De auteur	132



Toekomstbeelden

Wat zijn de belangrijkste data-vraagstukken voor de toekomst? Tussen de hoofdstukken door zijn bijdragen van experts opgenomen die met antwoorden en toekomstbeelden komen:

Dirk Helbing over een alternatieve digitale revolutie, pag. 22

Djeevan Schiferli over persoonsgebonden advies, pag. 24

Jeroen Dijkxhoorn over de toekomst van persoonsgegevens, pag. 26

Jeroen van den Hoven over Big Data en ethiek, pag. 43

Maurits Kreijveld over blockchain en prediction markets, pag. 46

Emma AI over de toekomst van AI, pag. 48

Mireille Hildebrandt over persoonlijke digitale assistenten, pag. 70

Eric van Tol over Big Data en de dienstensector, pag. 72

Rogier Creemers over het sociale kredietsysteem in China, pag. 74

Maurits Martijn over de dag dat Google valt, pag. 91

Theo de Vries over de onomkeerbaarheid van ICT, pag. 93

Marijn Janssen over Big Data en de maakbare samenleving, pag. 110

Brenno de Winter over Big Data als privacykans, pag. 112

VOORWOORD

Bij nagenoeg alles wat we doen genereren we data. Of het nu gaat over een bezoek aan de dokter, boodschappen doen of een reis plannen en maken, bij al deze activiteiten worden data gegenereerd die vervolgens ook weer ergens worden opgeslagen. Al deze data brengen ons nieuwe mogelijkheden in de vorm van diensten en services, maar er ontstaan ook ethische dilemma's rond eigenaarschap en gebruiksrecht. Zo kan het vrij beschikbaar maken van medische dossiers helpen om resultaten van klinische studies te vergelijken, maar wanneer verzekeraars hier ook toegang toe hebben kan dat leiden tot selectief uitsluiten van mensen of duur maken van bepaalde verzekeringen. Daarnaast geeft het bezit van data een zekere manipulatieve macht die zich uiteindelijk ook kan uiten in de vorm van geopolitieke spanningen. Er zijn in dit opzicht voldoende voorbeelden van regeringen die inzicht in (internet-)data van burgers gebruiken om bepaalde bevolkingsgroepen te controleren en te onderdrukken.

Op dit moment worden data vooral ingezet om gevraagd en ongevraagd op de persoon toegespitste proposities te doen. Dat kan variëren van slimme weekaanbiedingen van de supermarkt tot reisaanbiedingen die in de vorm van pop-ups als lastige horzels je beeldscherm blijven bevolken. Je kunt dit meestal met enige moeite uitzetten, maar de balans tussen nut en nuisance is niet voor iedereen hetzelfde. Er zijn ook voorbeelden van websites waarop patiënten (maar soms ook artsen) slim doktertje spelen door gegevens te delen en daarmee tot een snelle en betere diagnose te komen. Door gebruik te maken van de kennis en ervaringen van velen biedt Big Data zeker aanvullende mogelijkheden ten opzichte van de huidige medische praktijk. Dit is de situatie van vandaag; de grote vraag is echter waar het in de toekomst naartoe gaat met onze data, ofwel "data, quo vadis?"

In deze toekomstverkenning hadden we te maken met een lastig dilemma: gaat het nu om de data als zodanig of



zijn het juist de handige bewerkingen die maken dat deze data waardevol zijn? Mijn persoonlijke beeld hierbij is toch vooral dat data gezien kunnen worden als de noodzakelijke brandstof voor allerlei slimme bewerkingen. De eerste echte grote stappen die we hierbij zien hebben te maken met *deep learning* en *artificial intelligence*. In het geval van deep learning zijn eerder gegenereerde *real-life data* de toets voor een systeem. Door het kunstmatig gegenereerde antwoord steeds te vergelijken met de werkelijkheid treedt een leereffect op waarmee het vervolgens mogelijk is om steeds betere antwoorden of voorspellingen te geven. Ook kan hiermee steeds beter worden ingespeeld op een situatie-specifieke context. Dit alles maakt het systeem steeds ‘intelligenter’.

Waar gaan we uiteindelijk naartoe? Hierover lopen de meningen nogal uiteen. Zelf zie ik data als het vehikel waarmee we uiteindelijk de grens tussen mens en machine zullen slechten. Technieken als deep learning zijn hiervoor cruciaal, maar ook de verbinding tussen neurowetenschappen, high-techsystemen en (nano)materialen zal hiervoor van groot belang zijn. Wanneer we ons brein draadloos zouden kunnen koppelen met datasystemen en informatie uit ons brein zowel kunnen zenden als ontvangen, levert dit ongekende mogelijkheden op. Het zal duidelijk zijn dat de ethische dilemma’s die we nu rond data zien alleen maar heftiger naar voren zullen komen. Zo verwacht ik dat een belangrijk dilemma zal ontstaan rond de vraag op welk gebied systemen beslissingen zullen gaan overnemen van de mens; het besturen van een auto of een vliegtuig is nog wel te overzien, maar het bepalen van een medisch behandelplan (of misschien wel juist het stoppen daarvan) door een artificieel systeem zal al minder snel worden geaccepteerd.

In deze toekomstverkenning worden een aantal beelden geschetst over hoe de wereld er ten aanzien van data en daaraan gerelateerde zaken over 25 jaar uit zou kunnen zien. Backcasten van deze toekomstbeelden naar het nu maakt het voor onderzoekers en beleidsmakers mogelijk langetermijnsscenario’s te ontwikkelen om ons daarmee tijdig voor te bereiden op een sterk veranderende toekomst.

Ik wens u heel veel leesplezier, waarbij ik ervan overtuigd ben dat het voorliggende boekwerk u zal aansporen tot creatief denkwerk.

Jos Keurentjes

Chief Scientific Officer, en lid van Raad van Bestuur TNO

SAMENVATTING

We leven in tijden van digitale overvloed en de digitalisering van onze wereld lijkt geenszins af te nemen. De hoeveelheid data op aarde verdubbelt zich nu iedere tweeëneenhalf jaar en de snelheid waarmee dit gebeurt neemt toe. Binnen tien jaar zijn er naar verwachting zo'n 150 miljard sensoren via netwerken verbonden en de hoeveelheid data die hiermee geproduceerd wordt komt met nieuwe ambities. Nieuwe technologie brengt de mogelijkheid om alles wat we zien, doen, horen, denken en uitspreken te digitaliseren. Het antwoord op de vraag 'wat kunnen we niet in data vatten?', moet onder voortschrijdend inzicht steeds worden aangepast.

Deze publicatie presenteert de uitkomsten van een toekomstverkenning naar de sociale impact van Big Data. Aan de hand van de resultaten van interdisciplinaire foresight-workshops, tientallen interviews met experts en externe bijdragen van dertien vooraanstaande denkers schetst de publicatie een palet aan toekomstbeelden met belangrijke datavraagstukken voor morgen, maar ook voor vandaag. Uit de toekomstbeelden en scenario's komen een aantal belangrijke lessen naar voren:

Kennisbescheidenheid. Met de opkomst van datatechnologie die rekenkracht, opslag en algoritmische nauwkeurigheid maximaliseert, krijgen we nieuwe gereedschappen om naar de wereld te kijken en daarmee een goudmijn aan kansen om tot meer kennis te komen. Toch is er een

noodzaak voor kennisbescheidenheid. Zelfs als we met duizenden data-elementen – waaronder genetische informatie, locatiedata en een goed gevuld profiel – conclusies willen trekken over mensen moeten we beseffen dat de analyse onjuist kan zijn en niet blind vertrouwen op data-technologie. Big Data verleidt ons om te denken dat we een objectief oordeel over een persoon kunnen vellen. Het is in een tijdperk van Big Data belangrijk de morele autonomie van de mens te beschermen en mensen, ongeacht de data, eigen keuzes te laten maken en hun vrijheid te bewaken.

Mens centraal. "Als je niet betaalt voor een dienst ben jij niet de consument maar het product", stelde Andrew Lewis. Een kritiekpunt in onze relatie met organisaties die data over ons vastleggen is dat men niet goed begrijpt dat onze data verhandeld worden, hoe dit gebeurt en wat de waarde van onze data is. De in deze publicatie gepresenteerde machtsscenario's presenteren alternatieve datamodellen die laten zien dat decentralisatie van datadiensten zowel materiële als immateriële waarde zou kunnen opleveren voor gebruikers en de samenleving.

Nederland, dataland? Datagedreven technologie kan een belangrijke motor zijn voor economische groei en sociale vooruitgang. Hiervoor zullen het bedrijfsleven, het maatschappelijk middenveld, de overheid en de wetenschap moeten

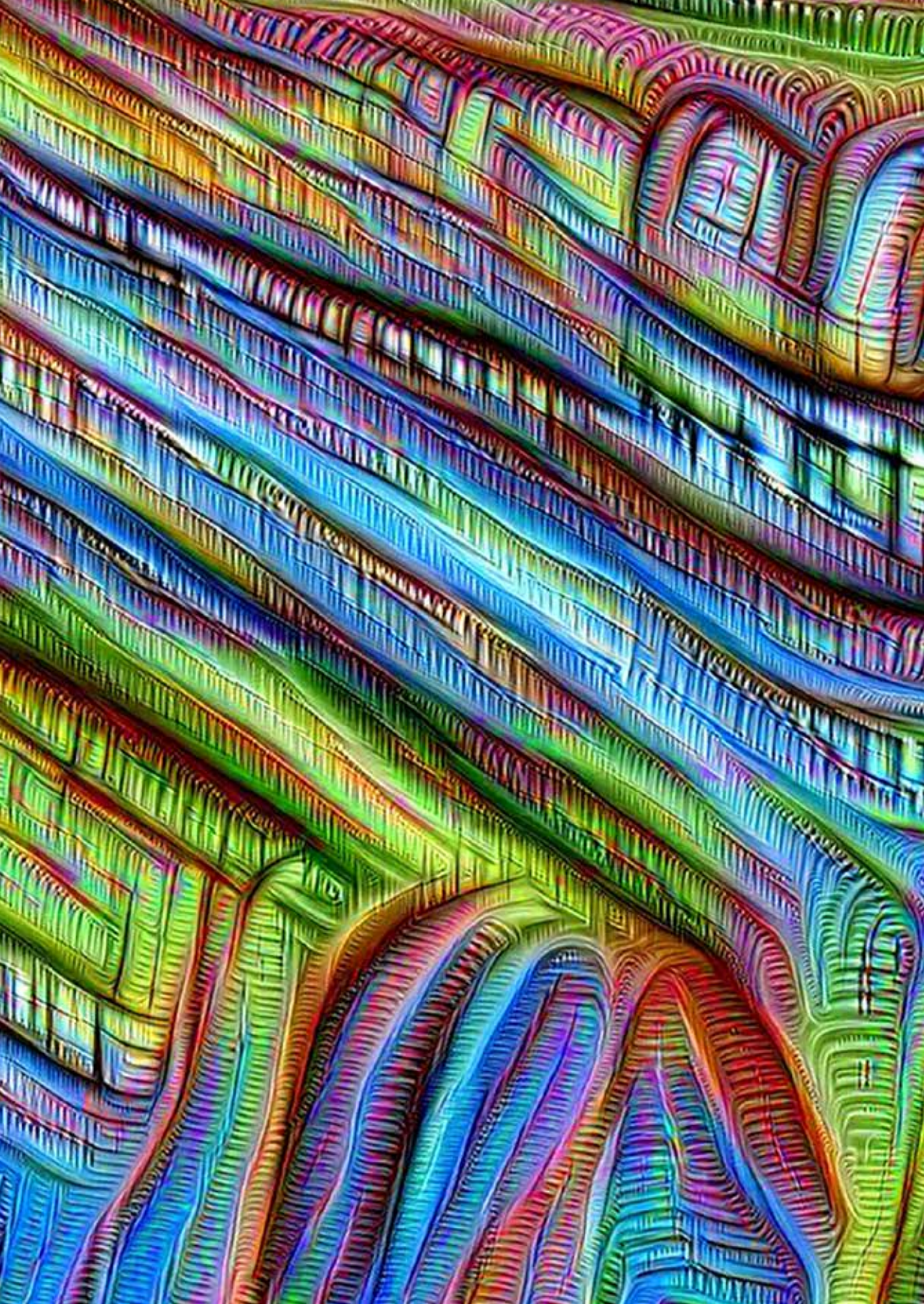
samenwerken om zowel aandacht voor technologie als gepaarde risico's te hebben. De spanning die nu ervaren wordt tussen Big Data à la Silicon Valley en onze eigen grondrechten zou een uitgangspunt kunnen bieden om lokaal tot nieuwe infrastructuur, data-applicaties, en intelligente systemen te komen die lokale waarden en de bescherming van de mens en eigen data centraal stellen.

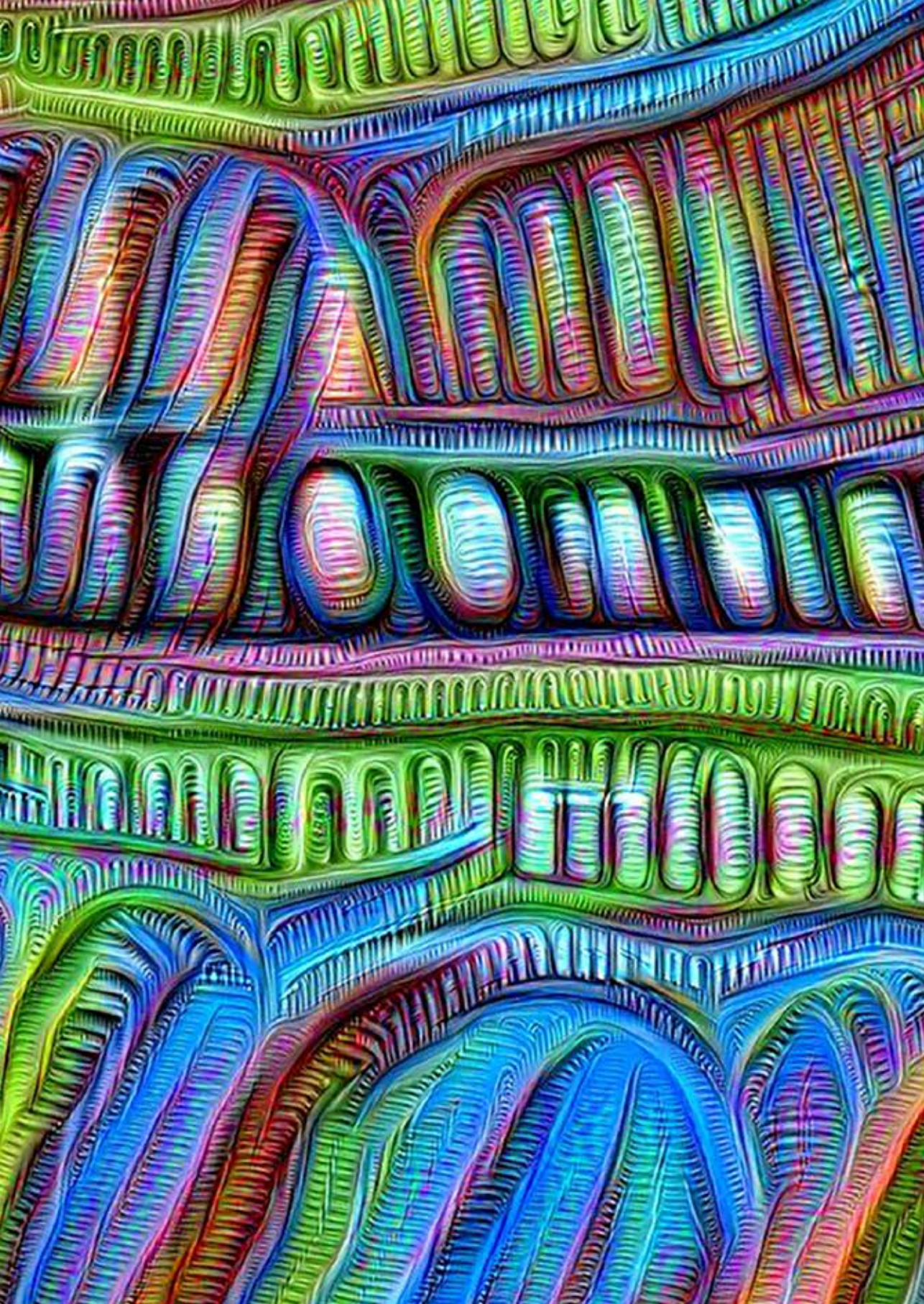
Big Data en pluraliteit. In het Big Data-veld is meer behoefte aan meerstemmigheid. Niet alleen in de discussie over het fenomeen, maar ook in de ontwerpfase, op bestuursniveau, in het aanstellen van ethische commissies is diversiteit wenselijk. Als we dit niet doen riskeren we dat machines, misschien wel heel weldenkende machines, een afspiegeling worden van een klein gedeelte van de samenleving en bij anderen juist weerstand zullen oproepen. We zouden juist het pluralisme, wat we in onze samenleving als belangrijke waarde bestempelen, in onze systemen moeten inbouwen. We zijn gebaat bij de wisdom of crowds, ook als de crowd voor een gedeelte uit slimme systemen bestaat.

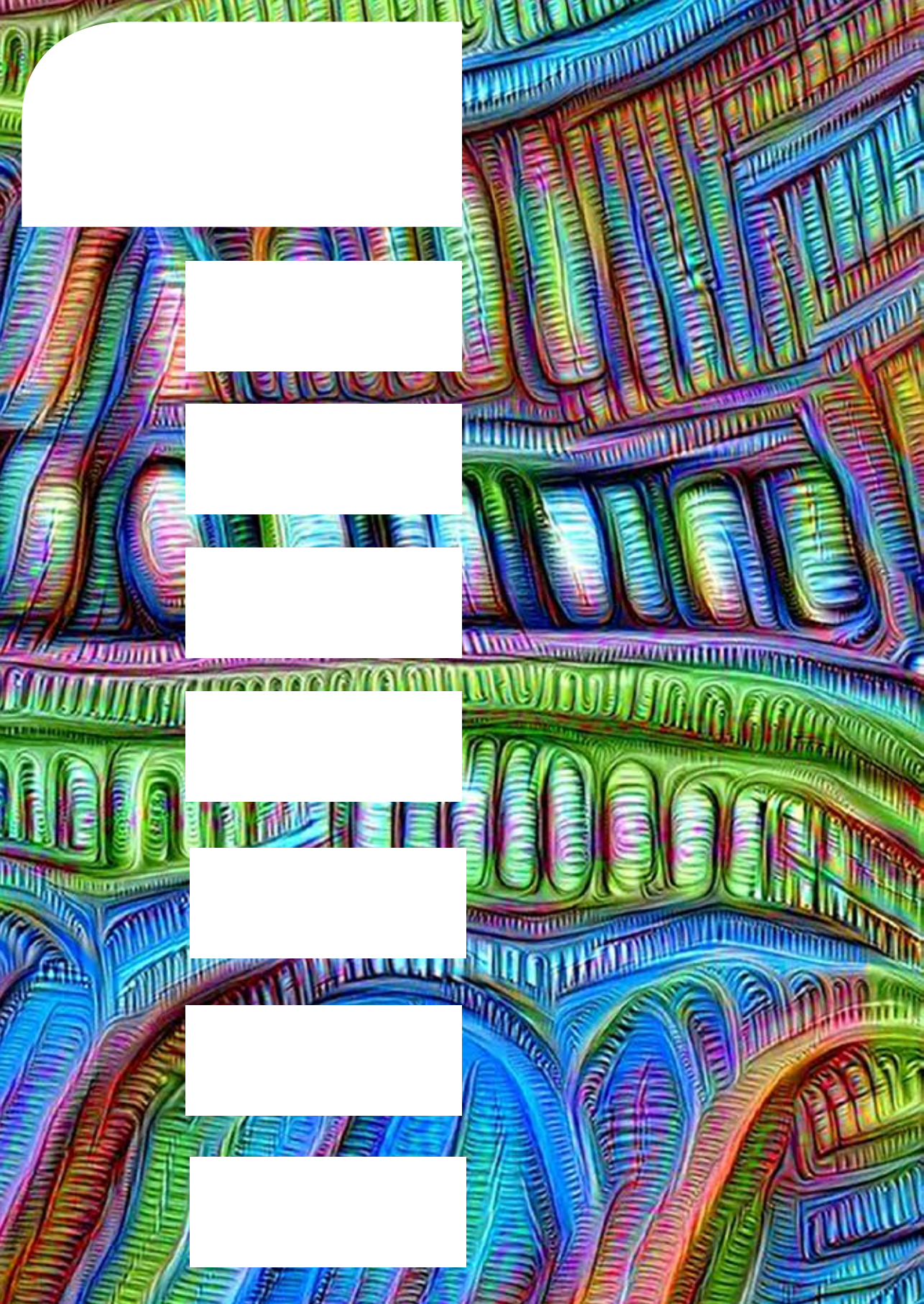
Bewust van onbewuste intelligentie. Hoewel de realisatie van superintelligentie volgens de meeste KI-experts nog toekomstmuziek is, zien we de komende jaren overal om ons heen nieuwe intelligentie opduiken. Sprekende huiskamers, denkende straatverlichting en persoonlijke

digitale assistenten staan in de virtuele rij om u van dienst te zijn. Intelligente zelflerende systemen zijn op komst en gebruikers en zelfs bedenkers van dit soort systemen zullen steeds minder grip hebben op de output hiervan. We hebben te maken met steeds autonomer wordende intelligente machines en moeten leren anticiperen op interactie met deze non-organische intelligentie.

Big Data schept hoge verwachtingen voor vele domeinen; van de zorg tot landbouw, onderwijs tot milieu, van de financiële sector tot transport. Hoewel er veel over Big Data geschreven is en veel organisaties de mogelijkheden en beperkingen verkennen ontbreekt er, vooral in de Nederlandse context, een verkenning naar de impact van deze technologie op de lange termijn. Wat ontbreekt zijn alternatieve beelden van onze samenleving, onze organisaties en onze overheden waarin verkend wordt hoe ze aan de hand van Big Data-technologie op geheel andere wijze zouden kunnen werken of worden ingedeeld. Welke bestuurstaken nemen algoritmes in de toekomst over? Zou persoonlijke informatie persoonlijk eigendom kunnen worden waaraan individuen kunnen verdienen? Hoe definiëren we privacy in 2040? In het beantwoorden van dergelijke vragen en het creëren van toekomstbeelden die ons alternatieve toekomsten helpen verbeelden ligt de kracht en meerwaarde van deze publicatie.







1. DATA ALS DNA

We bouwen aan een nieuw universum. Bit voor bit digitaliseren we onze wereld en slaan deze op in virtuele ruimte. Volgens de Italiaanse schrijver Italo Calvino werken we aan 'het geheugen van de wereld', een volledig digitale kopie van onze fysieke wereld. En we gaan verder dan kopiëren. We maken modellen van onze wereld waarin we simuleren hoe het weer is, hoe het was en hoe het zal zijn. We maken bedrijfsmodellen om grip te krijgen op logistiek, orders en productie. Financiële modellen helpen ons bij het maken van investeringen. Misdaadmodellen voorspellen waar een volgende overval zal plaatsvinden. We scheppen digitale werelden met het doel om meer grip te krijgen op onze eigen werkelijkheid. En onze eigen werkelijkheid wordt steeds digitaal. Een steeds groter deel van onze communicatie is digitaal, we spenderen meer tijd online, onze huishoudens en bedrijven zijn virtueel verbonden en ons digitale alterego wordt actief of passief samengesteld.

In honderden databases, waarvan u de meerderheid waarschijnlijk niet kent, is informatie over u opgeslagen. Uw mailbox archiveert een belangrijk gedeelte van uw correspondentie en contacten, veelal met voornaam, achternaam, e-mailadres, telefoonnummer en beroep. Locatiedata toont waar u de afgelopen maanden van dag tot dag bent geweest. Ongemerkt produceert u een lange lijst van zoekresultaten en browsegeschiedenis, vaak tientallen registraties per dag, die inzicht geven in uw voorkeuren, meningen, kennis, uw carrière en sociale contacten. De data is persoonlijk:

Uw volledige naam. Huisadres. Foto's. Video's. Emailadressen. BSN-nummer. Paspoortnummer. IP-adressen. Kentekenbewijs. Gedigitaliseerd profiel van uw gezicht, vingerafdrukken, schrift. Creditcardnummers. Geboorteplaats. Genetische Informatie. Telefoonnummers. Inlognamen. Leeftijd. Geboortedatum. Naam van uw werkgever. Uw professionele geschiedenis. Uw medische dossiers. Geslacht. Etnische achtergrond. Uw strafblad. Cookies. Contacten. Status-updates. Gedachten. Gevoelens.

Big Data-toepassingen zorgen ervoor dat deze lijst zich constant uitbreidt. Aan de hand van deze informatie, krachtige algoritmes en snelle datasystemen worden nieuwe verhalen over u gemaakt. Bent u geneigd een product te kopen, heeft u een groot sociaal netwerk, bent u geschikt voor een bepaalde baan en hoe communiceert u? Data worden

steeds meer gezien als een blauwdruk van ons gedrag, als de behavioristische variant op DNA. Als je genoeg data hebt kun je inzicht krijgen in de manier waarop een organisatie functioneert of een mens handelt, spreekt, slaapt of stemt. Je kunt met data in kaart brengen wat de geschiedenis is van iemands doen en laten en als je voldoende weet kun je een poging doen de toekomst te voorspellen. Deze informatie heeft invloed op onze levens: we worden aan de hand van data-analyses gefouilleerd op luchthavens, komen in aanmerking voor korting op bepaalde producten, krijgen banen aangeboden of worden versierd op het net.

Het proces is in zekere zin onomkeerbaar. Zelfs als u uw e-mailaccounts nu zou sluiten, uw sociale-media profielen zou opzeggen en van het digitale netwerk zou verdwijnen, blijft u nog jaren bestaan in de digitale ruimte.

Meer data

De digitale ruimte dijt uit en doet dit met toenemende snelheid. Er gaat per seconde meer informatie over het internet dan twintig jaar geleden op het hele net te vinden was. Per minuut worden er meer dan 150 miljoen e-mails verzonden en 300 uur aan video-materiaal op YouTube gezet. Naar schatting hebben we meer bits opgeslagen dan er sterren zijn in het fysieke heelal.

Er is zoveel data dat we niet exact weten hoeveel het is. Velen hebben geprobeerd om een getal te geven aan de hoeveelheid informatie die bestaat. Martin Hilbert van de University of South California is hiervan een van de meest vooraanstaande onderzoekers. Hilbert schat dat we gedurende de afgelopen dertig jaar iedere tweeënehalf jaar een verdubbeling van digitale informatie hebben gezien. Dit komt neer op een totaal aantal van 10 zettabyte in 2016, waarbij een zettabyte in cijfers wordt uitgedrukt als 10^{21} ofwel 1.000.000.000.000.000.000 bytes (Hilbert 2016). Deze hoeveelheid is moeilijk te bevatten. Als het op dvd's zou worden gezet en we zouden die dvd's stapelen dan zouden we vijf stapels krijgen die allemaal tot de maan reiken.

Megabyte	1.000.000 bytes
Gigabyte	1.000.000.000 bytes
Terabyte	1.000.000.000.000 bytes
Petabyte	1.000.000.000.000.000 bytes
Exabyte	1.000.000.000.000.000.000 bytes
Zettabyte	1.000.000.000.000.000.000.000 bytes
Yottabyte	1.000.000.000.000.000.000.000.000 bytes

We leven in tijden van digitale overvloed en de digitalisering van onze wereld lijkt geenszins af te nemen. De Britse onderzoekers Mayer en Schonberg vergelijken wat nu gebeurt met de print-revolutie die volgde op het uitvinden van de drukpers in Gutenberg. Van 1453 tot 1503 werden ongeveer acht miljoen boeken geprint. Dit stond ongeveer gelijk aan het aantal geschreven teksten die in de 1200 jaar daarvoor in Europa waren geproduceerd. Met andere woorden, tijdens de print-revolutie kostte het zo'n vijftig jaar om de bestaande informatie te verdubbelen. Nu gebeurt dit elke 2.5 jaar en neemt de snelheid nog altijd toe. Binnenkort zullen de dingen om ons heen, zoals onze koelkasten en kleding, ook verbonden worden met het internet.

Dirk Helbing en Jeroen van den Hoven van de TU Delft schatten dat binnen tien jaar 150 miljard sensoren – twintig keer zoveel als er mensen zijn op aarde – via netwerken worden verbonden. Dan verdubbelt de data die we produceren ieder etmaal (Helbing *et al.* 2016), wat betekent dat we dan in twaalf uur tijd evenveel data produceren als we tot dat moment hebben gedaan. Deze data komt met nieuwe ambities. Nieuwe technologie brengt de mogelijkheid om alles wat we zien, doen, horen en uitspreken te digitaliseren binnen handbereik. Het antwoord op de vraag 'wat kunnen we niet in data vatten?' moet onder voortschrijdend inzicht steeds worden aangepast.

Meer data, maar waarvoor?

"De meest waardevolle handelswaar die ik ken is informatie", stelde Gordon Gekko in de film *Wallstreet* (1987). Aan de hand van Big Data krijgen we meer informatie over – en dus meer grip op – de wereld en haar toekomst. Elke keer dat jij informatie zoekt op internet, zoekt het internet ook naar informatie over jou. Zo wordt weleens provocerend gezegd dat het internet je beter kent dan jijzelf.

Deze ontwikkelingen leiden tot een breed scala aan innovatieve toepassingen. Bij politiekorpsen over de gehele wereld wordt *predictive policing* ingezet, waarbij op basis van een grote dataset met misdaadhistorie, plaats, tijd, weersomstandigheden en andere factoren voorspeld wordt waar en wanneer misdaad zal plaatsvinden (voorbeelden hiervan zijn het Criminaliteit Anticipatie Systeem in Nederland, of Predpol internationaal). Aan de hand van dit soort data bepaalt het systeem welke routes agenten moeten rijden en worden mogelijke daders en slachtoffers vroegtijdig geïdentificeerd. In Los Angeles zijn door middel van dergelijke technieken inbraken met 33% en geweldsdelicten met 21% afgenomen (Mohler *et al.* 2015).

Erik Brynjolfsson van MIT schat dat negentig procent van alles wat bedrijven doen inmiddels digitaal is of digitaal traceerbaar is. In de jaren '90 was dit nog niet eens twintig procent. Warenhuizen experimenteren met het aanpassen van hun reclame en de prijzen van hun miljoenen producten op basis van datasets met klantengedrag, aankoopgeschiedenis, prijzen van concurrenten, twitter-berichten, en een media-analyse.

Via Google's Flu Trends worden voorspellingen gedaan over het uitbreken van epidemieën op basis van correlaties tussen relevante zoektermen. In de zorg worden systemen ontwikkeld die middels gestructureerde data van inspecteurs en ongestructureerde data zoals berichten op sociale media en blogs, risico's bij zorginstellingen signaleren. Het idee is dat dit tot kennisgestuurd toezicht leidt waarbij toezicht niet meer simpelweg een overheidstaak is, maar ook burgers betreft door hun digitale uitingen mee te wegen.

Datasystemen worden bij dit soort taken steeds meer van kunstmatige intelligentie (KI) voorzien zodat nieuwe vormen van mens-machine-interactie mogelijk zijn. Er worden allerlei applicaties ontwikkeld die gebruikers op basis van data-analyses advies kunnen geven en zelfs als persoonlijke assistent kunnen fungeren. Naarmate de technologie zich ontwikkelt zien we dat systemen die kunstmatige intelligentie en grote datasets combineren niet alleen maar vragen kunnen beantwoorden en aanbevelingen kunnen doen, maar ook autonome beslissingen kunnen nemen en uitvoeren. Dit betekent dat we met steeds intelligentere systemen te maken krijgen en daar een weg in moeten vinden. Hoe denken datasystemen? Wat kunnen we machines toevertrouwen? Worden ze concurrenten van ons?

Omdat datagedreven technologie een hogere handelingssnelheid heeft dan mensen, kunnen bepaalde systemen op een bovenmenselijk niveau opereren. Zo veranderen zelfsturende voertuigen in toenemende mate de transportsector, wordt in de financiële sector ongeveer zeventig procent van alle markttransacties uitgevoerd door zelfhandelende algoritmes, en worden algoritmes op de beurs als waakhonden ingezet om real-time naar verdachte transacties en anomalieën te speuren (Helbing 2015).

De verdere ontwikkeling van Big Data-technologie zal zodoende leiden tot bepaalde verschuivingen in de samenleving en verandert de manier waarop overheid, bedrijfsleven en wetenschappen zich tot elkaar en tot de samenleving verhouden. Bij de inzet van deze technologieën zullen organisaties, overheden en de wetenschappen worstelen met vraagstukken over de kwaliteit en toepassing van data (is Big Data ook betere data?), het beheer ervan (van wie zijn de data?), en maatschappelijke vraagstukken (welke impact heeft Big Data op onze machtsverhoudingen, rechten en identiteit?).

Nieuwe profielen

Een Big Data-casus ter illustratie. Aan de hand van nieuwe data kunnen meer uitgebreide en betere profielen over jou worden opgesteld. Je kunt deze profielen zelf beheren door op allerlei sociale netwerken te bepalen wat anderen over jou te zien krijgen, maar bedrijven en overheden werken tegelijkertijd ook aan hun eigen profielen over jou. Tijdens dit onderzoek ontmoetten we Crystal. Crystal is geen vrouw, zoals haar naam doet vermoeden, het is geen man, het is zelfs geen mens. Crystal is een set algoritmes

die het web afspeurt naar toegankelijke data over personen en op basis hiervan adviseert hoe je het beste met deze persoon kunt communiceren. Als een mail wordt opgesteld aan bijvoorbeeld Mark Rutte stelt Crystal dat je te maken hebt met een ‘snelle denker, ambitieus, overtuigend, neemt berekend risico’. Wat voor Rutte als natuurlijke communicatie geldt, somt ze als volgt op:

 **It comes naturally to Mark to...**

- Cut a conversation short to get something done**  
- Speak very directly or bluntly**  
- Patiently wait in line**  
- Be frustrated if someone is late to a meeting**  

Crystal komt in de vorm van een plug-in voor de mail-applicatie en adviseert tijdens het typen van een mail wat te schrijven. Voor Mark Rutte adviseert Crystal: “begin met een non-essentiële vraag, zoals ‘hoe was je vakantie?’; Schrijf niet meer dan 3 regels, gebruik afkortingen en een informele stijl.” Ondertussen stelt ze voor om bepaalde woorden te schrappen en anderen te herformuleren.

Crystal is imperfect, maar roept interessante, typische Big Data vragen op. Welke data worden gebruikt om tot dit profiel te komen? Zijn al die data wel geverifieerd? Telt een bericht op de website van de overheid zwaarder mee dan informatie op bijvoorbeeld Mark Rutttes facebook-pagina? Wie is de eigenaar van dit profiel? Maar ook stelt het de vraag of menselijke communicatie zal veranderen aan de hand van Big Data-analyses. Stel dat ik gebruik maak van een persoonlijke digitale assistent om te communiceren en mijn gesprekspartner gebruikt die ook: wordt onze communicatie dan veel efficiënter en gemakkelijker? Of verliezen wij een stukje eigenheid door deze sturende datafilters?

In situaties waarbij men met een onbekende te maken heeft, biedt Crystal een uitkomst. Bedrijven die producten aan de man willen brengen kunnen beter bepalen welke potentiële klanten ze het beste kunnen aanschrijven en ook hoe ze dit het meest effectief kunnen doen. Headhunters kunnen zo op voorhand al een selectie maken van personen die geschikt zijn voor hun opdrachtgever en beter bepalen of iemand geschikt is voor een bepaalde functie en bedrijfssetting. Overheden kunnen op een meer persoonsgerichte manier communiceren met burgers en mensen met gelijke interesses kunnen snel aan elkaar gekoppeld worden. Als de data kloppen tenminste.

Zoals Frank Pasquale aangeeft in zijn boek *Black Box Society*, leven we in een samenleving die steeds meer vertrouwen heeft in datasystemen die feitelijk als een zwarte doos werken (Pasquale 2015). Voor de eindgebruiker is het onduidelijk welke data er in het systeem zitten, of de data juist zijn, en wat ermee gedaan wordt. Voor Crystal is het niet anders: op haar website (www.crystalknows.com) is terug te vinden dat ze alle publieke bronnen gebruikt die ze kan vinden, maar het is in individuele gevallen niet duidelijk welke dit exact zijn. Voor de afnemer van de dienst is het bij Crystal, maar ook bij sites als Google, niet duidelijk hoe bepaalde bronnen gerangschikt worden of hoe ze gewogen worden ten opzichte van elkaar. Is de informatie op LinkedIn meer accuraat dan die op de site van je werkgever? Het is eveneens onduidelijk welke stukjes data leiden tot welke inzichten. Als ik bij hobby's invul dat ik een interesse heb in filosofie, betekent dit dan dat ik geschikt dan wel ongeschikt ben voor een bepaald type baan?

Wat duidelijk is, is dat algoritmes die data zoeken, wegen, filteren en presenteren steeds bepalender zijn voor allerlei aspecten van ons leven. Zo vroeg de Amerikaanse voormalige presidentskandidaat Rick Santorum Google te interveniëren omdat hij vond dat hun algoritmes een negatieve impact hadden op een van de belangrijkste processen in de samenleving, de democratie (2011). Zijn achternaam was immers door duizenden internet-activisten gelinkt aan vulgaire handelingen en wat hem dwars zat was dat de nieuwe betekenis van zijn naam het eerste is wat Google toonde als je erop zocht. Omdat veel mensen gebruikmaken van de zoekmachine en over het algemeen slechts de eerste paar zoekresultaten bekijken, kan een dergelijk negatief bericht bovenaan de lijst leiden tot beïnvloeding van het stemproces, redeneerde de Republikein. Een aantal maanden later paste het bedrijf haar algoritmes aan om manipulatie van zoekopdrachten te voorkomen, maar stelde tevens dat er geen handmatige aanpassing van resultaten plaatsvond.

Omdat het onduidelijk is hoe de automatische aanpassing van resultaten precies werkt en er geen inzicht in wordt gegeven is hiermee het vraagstuk nog niet geheel afgedaan. De vraag blijft hoe groot de invloed van datasystemen is op onze democratie en op onze levens. Richting de toekomst is de vraag welke impact data de komende jaren hebben op u als burger, als consument en als mens? Dit vraagstuk staat in dit boek centraal.

Meerwaarde toekomstverkenning

Big Data maakt beloftes in vele domeinen; van de zorg tot landbouw, onderwijs tot milieu, van de financiële sector tot transport. De technologie achter de toepassingen is echter vergelijkbaar en om die reden onderzoekt dit rapport niet één sector, maar richt het zich op Big Data-technologie in algemene zin. Er is veel over Big Data geschreven en veel organisaties verkennen de mogelijkheden en beperkingen van Big Data. Wat

ontbreekt, vooral in de Nederlandse context, is een verkenning naar de impact die deze technologie op de lange termijn heeft (15-25 jaar vooruit). Wat ontbreekt zijn alternatieve beelden van onze samenleving, onze organisaties en onze overheden waarin verkend wordt hoe ze aan de hand van Big Data-technologie op geheel andere wijze zouden kunnen werken of worden ingedeeld. Welke bestuurstaken nemen algoritmes in 2030 over? Kunnen we Big Data-technologie gebruiken om menselijke intelligentie te verbeteren? Zou persoonlijke informatie persoonlijk eigendom kunnen worden waaraan individuen kunnen verdienen? Hoe definiëren we privacy in 2040?

In het beantwoorden van dergelijke vragen en het creëren van toekomstbeelden die ons alternatieve toekomsten helpen verbeelden ligt de kracht en meerwaarde van deze publicatie. Het verkennen van langetermijntoewikkelingen vervult een belangrijke rol bij het in kaart brengen en voorstelbaar maken van de toekomst.

Bij de langetermijnhorizon hoort, zeker als het gaat om ontwikkelingen in het data-domein, een grote onzekerheid. Daarom baseert deze toekomstverkenning zich op wetenschappelijk onderzoek naar technologie en de mens, maar zijn we ook in de veronderstelling dat we 'wat als'-vragen moeten durven stellen. Maarten Hajer stelt dat we nieuwe verhalen nodig hebben om de samenleving perspectief te geven: "voor die verbeeldingskracht ben je bij de huidige wetenschap aan het verkeerde adres, die doet daar niet aan. De verbeelding is het domein van de kunst, van de creatieven, van de sciencefictionsschrijvers" (Postma 2016). De verkenning biedt daarom ook ruimte voor experimentele verhalen die een beeld schetsen van mogelijke toekomsten. Door na te denken over de toekomst en deze te verbeelden, proberen we met deze verkenning de lezer te inspireren en te provoceren en tegelijkertijd vraagstukken over de toekomst van Big Data in het hier en nu te agenderen.



Leeswijzer

Na dit inleidende hoofdstuk richt hoofdstuk 2 zich op de technologische trends in het Big Data-domein. Wat zijn de belangrijkste technologische ontwikkelingen op het gebied van digitalisering en wat kunnen we verwachten op basis van zichtbare trends en signalen?

Hoofdstuk 3 gaat in op de sociale gevolgen van datagedreven technologie. Hoe verandert Big Data de manier waarop we ons verhouden tot elkaar, tot bedrijven en tot de overheid? Datagedreven technologie brengt spanningsvelden met zich mee waar we ons bewust van moeten zijn. Leidt meer data tot meer service of juist tot meer surveillance? Met het oog op de toekomst gaat dit hoofdstuk in op Big Data, transparantie, identiteit en machtsverhoudingen.

Hoofdstuk 4 en 5 nemen een voorschot op de toekomst en presenteren scenario's die gebaseerd zijn op de interviews en workshops die gehouden zijn in het kader van deze verkenning. De scenario's beschrijven toekomsten met alternatieve datamodellen en gaan in op vragen over data-eigendom en databeheer.

In het laatste hoofdstuk trekken we conclusies uit de toekomstbeelden en staan we stil bij de vraag wat we nu kunnen en moeten doen.

Daarnaast zijn door het gehele boek toekomstbeelden opgetekend. Deze zijn geschreven door dertien denkers. Zo gaat Mireille Hildebrandt in op de rol van persoonlijke digitale assistenten in de toekomst, spreekt Jeroen van den Hoven over Big Data en ethiek en waarschuwt Dirk Helbing voor de dictatuur van *singularity*. Marijn Janssen schrijft over Big Data en het ideaal van een maakbare samenleving en Theo de Vries anticipeert onder andere op het expertisetekort dat in Nederland dreigt te ontstaan op het gebied van ICT. Maurits Martijn schrijft over de dag dat Google valt en Rogier Creemers trekt lessen uit het sociale kredietsysteem dat China optuigt om al haar burgers via Big Data te monitoren. Jeroen Dijkhoorn roept op tot nieuwe applicaties die de privacy van persoonsgegevens waarborgen en Eric van Tol schrijft vanuit 2036 over Big Data en de dienstensector. Djeevan Schiferli reflecteert op de functionaliteit van persoonsgebonden advies, Maurits Kreijveld heeft het over de voorspellende waarde van blockchaintechnologie en de verbetering daarvan door prediction markets en Brenno de Winter schrijft over de kansen die Big Data biedt om privacy beter te waarborgen. Emma (een KI-systeem) schrijft een stuk over haar eigen toekomst, over de toekomst van kunstmatige intelligentie dus.

Bouwen aan een bloeiende wereld

Klimaatverandering, financiële crises, massamigratie, terrorisme, ongelijkheid; we hebben een aantal existentiële problemen op deze planeet. Geen van deze problemen hebben we nog kunnen oplossen met ICT, en ik vraag me af of we ze zelfs hebben kunnen terugdringen.

Ondanks onze beschikking over uitstekende technologie en meer data dan ooit tevoren, zetten we ze klaarblijkelijk niet op de juiste manier in. Futuristen denken dat we een omslagpunt, de zogenaamde ‘singulariteit’, bereiken als we onze technologie gewoon blijven doorontwikkelen. Sommigen stellen dat dit over vijf of vijftien jaar vanaf nu zal gebeuren. IT-systemen zouden dan zo ongelooflijk intelligent geworden zijn dat ze al onze problemen in één magische klap zouden oplossen. Dit is absoluut naïef, mensen die dit zeggen begrijpen niet wat de digitale revolutie betekent.

We zijn bezig om alles met elkaar te verbinden; data, netwerken, infrastructuur, dingen, mensen. Dit leidt tot een niveau van verbondenheid en complexiteit, dat door gecentraliseerde systemen niet kan worden begrepen en zeker niet op een top-down manier kan worden bestuurd. Niemand is slim genoeg om dit soort complexiteit te begrijpen, ook een toekomstig KI-systeem niet.

In plaats van het streven naar singulariteit moeten we onze waarden van diversiteit en pluraliteit bewaken en deze toepassen op technologie. We hebben een nieuw paradigma nodig dat hierop in kan spelen: een van zelfregulering en gedistribueerd bestuur. Het Internet of Things in combinatie met blockchaintechnologie en complexiteitswetenschap reiken ons de hand om dit voor elkaar te krijgen.

Collectieve intelligentie

In een kathedraal met complexe ornamenten is er niet één enkel perspectief dat de volledige fraaiheid kan vatten. Ditzelfde geldt voor elk complex fenomeen; we hebben verschillende perspectieven en collectieve intelligentie nodig om ze te doorgronden. Momenteel is het bij onze bestuursmodellen zo dat diegene met de meeste macht beslist welk perspectief gekozen wordt, terwijl andere perspectieven worden gemarginaliseerd. Daardoor krijgen we gesimplificeerde oplossingen die niet goed werken. We moeten afscheid nemen van het ‘51% beslist voor 100%-model’. Onze regeringen hebben kleine meerderheden en bepalen voor vier, acht of nog meer jaren wat er gebeurt. De rest moet maar afwachten. Deze aanpak streeft er niet naar alle perspectieven in de samenleving mee te wegen, maar is macht-georiënteerd.

Een systeem waarbij de vossen alle konijnen opeten, is uiteindelijk slecht voor de vossen zelf, omdat er niets te eten overblijft. We moeten dus werken aan een meer holistische aanpak. Het ergste wat we onze samenleving kunnen aandoen is oversimplificeren.

De filter bubbel

Big Data kan bijdragen aan dit soort simplificatie. Als we niet oppassen ontstaat er een matrix-achtig scenario, waarbij we de wereld uitsluitend door filters kunnen zien. Er ontstaat dan een innovatieprobleem, aangezien het moeilijker wordt om 'out of the box' te denken in een samenleving waarbij we constant *genudged* worden.

We hebben behoefte aan bestuursplatformen die collectieve intelligentie stimuleren, en die de verschillende perspectieven samen laten komen. Als je alleen efficiëntie en winst voor ogen hebt dan gooi je het kind met het badwater weg.

Alternatief digitaliseren

We hebben te lang verondersteld dat er maar één manier van digitalisering bestaat: de Silicon-Valley-route. Dat betekent gas geven en zo snel mogelijk innoveren. Maar net als hard rijden op een mistige dag kan exponentiële groei leiden tot brokken. Digitale transformatie kan op verschillende manieren worden ingezet. Japan, bijvoorbeeld, gebruikt hiervoor de techniek van *backcasting*. Zij stellen zichzelf de vraag: "waar willen we als samenleving over 20 jaar staan?" Daarna kijken ze welke technologie nodig is om daar te komen. Dit is een alternatieve aanpak, die wereldwijd een gunstig effect zou kunnen hebben. De meeste problemen op aarde zijn door de mens veroorzaakt; door systemen die wij zelf gecreëerd

hebben. Er zijn echter vele manieren om een samenleving aan te sturen, en we kunnen er zeker één vinden waarbij er geen mensen sterven ten gevolge van door de mens veroorzaakte tekorten.

Nieuwe waardering

Waarom zouden wij bijvoorbeeld in de toekomst nog gebruikmaken van een zo eenvoudig waarderingssysteem als ons huidige geldstelsel? Waarom innoveren we dit systeem, dat in wezen duizenden jaren oud en achterhaald is, niet? We hebben de mogelijkheid om een nieuw systeem uit te vinden dat meer multidimensionaal is, en rekening

Niemand is slim genoeg om dit soort complexiteit te begrijpen, ook een toekomstig KI-systeem niet.

houdt met externe prikkels en de wensen van velen.

Zo'n nieuw systeem moet gebaseerd zijn op het subsidiariteitsbeginsel. Dit betekent dat het heel decentraal geregeld wordt, waarbij prijzen kunnen verschillen per buurt. Het maken van geluid kan bijvoorbeeld in een rustige buurt heel duur zijn, terwijl het in de binnenstad een stuk goedkoper is. Een flexibele aanpassing aan plaatselijke waarden en behoeften is van wezenlijk belang.

Route voor de toekomst

Ik houd hier geen betoog tegen Big Data, of tegen kunstmatige intelligentie. Het is onze taak om de voordelen van technologie te maximaliseren en ongewenste effecten te minimaliseren. Er is geen technologie die geen gevaren met zich meebrengt. Wel is het de vraag waarom we niet meer tijd en middelen besteden om de digitale revolutie zo in te richten dat iedereen ervan kan profiteren.

Persoonlijke raad

Het is 25 november, een koude maar zonnige herfstmiddag in Den Haag.

Ik heb een lunchafspraak met een onbekende man. Hij is geïnteresseerd in mijn ervaringen met Big Data en analytics. Nu kan ik gemakkelijk hele dagen met dit soort afspraken vullen, meestal zijn het dan verkopers van analyse-software, of op zoek naar projecten, of recruiters die data-wetenschappers willen wegzetten. Het lijkt wel alsof tegenwoordig iedereen in de Big Data actief is. Maar deze man kwam op aanbeveling van een gewaardeerd collega, en je moet toch lunchen...

Het gesprek ging snel bergafwaarts. De eerste 5 minuten verliepen ongemakkelijk en gaven me een *unheimliches Gefühl*, zoals mijn Duitse collega's zouden zeggen. Het leek alsof hij me goed kende, maar we hadden elkaar nog nooit ontmoet. De vragen die hij stelde, de woorden die hij gebruikte, de onderwerpen waarmee hij kwam en de manier waarop hij mij aan het praten kreeg... Ik voelde me gemanipuleerd, maar was daar niet echt zeker van, wat het des te vervelender maakte.

Na 10 minuten legde hij uit dat hij mijn naam had ingevoerd in *CrystalKnows* en gaf me een vel papier waarop stond welke vragen hij mij op welke manier zou moeten

stellen, en welke woorden wel en niet te gebruiken. Later in het gesprek gebeurde het een paar keer dat ik iets wilde zeggen, maar mijn woorden inhield omdat het Crystal's profiel van mij zou bevestigen.

En dat wilde ik niet. Ik ben geen nummer, ik ben een vrij man! Iemand die probeert om zijn persoonlijke gegevens op het web enigszins onder controle te houden. Ik gebruik geen Facebook, en probeer Google en soortgelijke systemen te vermijden. Toch gaf CrystalKnows een 90% correctheid aan van haar gegevens over mij. En dat klopte best wel.

**Ik ben geen nummer,
ik ben een vrij man!**

Tegenwoordig heb ik te vaak dit soort ervaringen. Dat mijn smartphone (Apple, want iets meer controle over mijn privacy) me suggereert te vertrekken naar een afspraak. Heeft een app mij om toestemming daarvoor gevraagd? Wat wordt er nog meer gevolgd? Of hoe gemakkelijk het is om een mobiele telefoon zonder VPN-diensten te hacken, en over je schouder mee te kijken voor je user-id's en wachtwoorden. Of je verstrekt je e-mailadres in een hotel en krijgt de dag erna spammails van totaal andere bedrijven! Een groot

deel van de bevolking gelooft nog steeds dat gratis diensten en apps ook inderdaad gratis zijn, of ze vinden dat ze niets te verbergen hebben. Maar zij missen het hele punt van privacy en van transparantie over wie bevoegd is welke persoonlijke gegevens beschikbaar te stellen op welk moment. Waarom zou iemand heel veel geld mogen verdienen aan mijn privégegevens en, erger nog, daar keuzes voor mij op baseren zonder mij inzicht of invloed te geven of te laten delen in de opbrengsten? Mijn volgende ontmoeting zal dus anders zijn. Ik bereid me altijd al voor op nieuwe contacten door vooraf te googelen (DuckDuckGo), te kijken of ze een LinkedIn-profiel hebben en of ik iets kan vinden op Pipl. Maar vanaf nu voeg ik daar CrystalKnows aan toe. Dit hangt nauw samen met mijn werk; ik werk op dagelijkse basis met Big Data-experts en startups die ons benaderen met fascinerende oplossingen.

Maar hoe zit het met mensen die niet zoveel weten van dit Big Data-domein? Of van kunstmatige intelligentie en *machine learning*? Wie is er straks nog in staat om 'slimme' huizen te begrijpen en om te gaan met de complexiteit ervan? Huizen met talloze apparaten die via het Internet of Things verbonden zijn aan verschillende leveranciers, die op hun beurt steeds meer gebruik maken van KI-toepassingen zoals Alexa, Siri en Cortana? En die op die manier, en om voor de hand liggende redenen, toegang krijgen tot steeds meer van onze gegevens.

Wat zou er gebeuren als...?

Er zijn startups onderweg die dezelfde Big Data-mogelijkheden en KI proberen in te zetten vóór de consument en burger. In plaats van alleen KI-Persoonlijke Assistenten van grote techbedrijven in te zetten – die vooral proberen nog meer van je te weten te komen en daar geld aan te verdienen en je ook nog proberen te manipuleren tot bepaalde beslissingen – zullen nieuwe startups PA's en de bijbehorende analytics aanbieden. Die alternatieve PA's zullen niet alleen even goed advies leveren, maar je daarnaast vertellen wat je eigen profiel op het web is, en op basis van deze inzichten kunnen herkennen wanneer je wordt gemanipuleerd of hoeveel jouw data op dat moment waard is. Nieuwe oplossingen waarbij jouw data van jezelf blijft en niet achter je rug om wordt verkocht aan derde partijen.

Dan zou er een wereld kunnen ontstaan waarin bedrijven eerst je vertrouwen moeten verdienen en waarbij ze een meerwaarde voor de afnemer moeten creëren om hun producten en diensten te kunnen verkopen. Dat is een grote uitdaging voor degenen wiens huidige inkomsten komen uit het moneteriseren van persoonlijke gegevens van klanten of het verkopen daarvan aan derden. Het is bemoedigend om te zien dat meer en meer risicodragend kapitaal naar startups gaat met hun nieuwe benadering gericht op transparantie en het teruggeven van de controle in de handen van de consument.

Zou het niet geweldig – en ook logisch – zijn als Europese bedrijven deze nieuwe ontwikkeling zouden aanvoeren?

Wilt u uw Big Data met of zonder... persoonsgegevens?

De aandacht voor Big Data leidt ertoe dat bedrijven eindelijk tot het besef komen dat klantgegevens uiterst waardevol zijn. Dit betreft niet alleen demografische gegevens van klanten maar ook gegevens die het klantgedrag vastleggen en de automatische profilering die hierop gedaan kan worden middels bijvoorbeeld *machine Learning* technieken. Bijvoorbeeld om direct de best passende aanbieding te kunnen doen voor een volgende aankoop of om potentiële fraude te voorkomen voordat het gebeurt. Maar wat betreft de bijbehorende regulering van gebruik zijn we nog in het 'Wilde Westen' van Big Data-land. Een mooie tijd voor pioniers (en boeven?), maar ook een gevaarlijke tijd voor een argeloos en onbevangen individu. Regulering? Dat is toch iets wat we juist *niet* willen...? Wellicht, maar net zoals met de uitvinding van de auto de noodzaak ontstond van specifieke verkeersregels en verkeerslichten, zal er een balans gevonden moeten worden tussen de organisaties – met mogelijkheden om persoonsgegevens om te zetten in waarde – en de personen aan wie die waarde 'onttrokken' wordt. De *General Data Protection Regulation* van de Europese Unie die in 2018 in werking zal treden is daar een voorbode van. De boetes op het niet naleven zijn significant groot en ook het toezicht op naleving wordt beter georganiseerd.

Dit is een signaal van verdere bewustwording bij regulerende instanties en individuen dat aangeeft dat het eigenaarschap van en de zeggenschap over persoonsgegevens goed geregeld moet zijn. "Toch niet weer een privacydiscussie rondom Big Data" hoor ik u denken. Nee, in plaats van te kijken naar het aspect van privacy wil ik liever kijken naar het gevolg van de toemerkende waarde van persoonsgegevens.

Stel nu dat het besef van deze waarde zodanig wordt dat individuen een grotere controle willen hebben over het wel of niet verstrekken van hun gegevens voor een specifiek doel. Regulering, of beter gezegd: marktregulering, is een voorwaarde om dit mogelijk te maken. Iedereen kent wel de hedendaagse voorbeelden van websites die allerlei niet noodzakelijke gegevens vragen of de software-apps die ellenlange contractuele voorwaarden voorleggen zonder mogelijkheid tot inspraak. Momenteel is er geen mechanisme dat het individu in staat stelt om controle over zijn gegevens uit te oefenen zonder de transactie in z'n geheel te weigeren.

Wat als er in de toekomst brokersystemen zijn waarmee het individu zelf de waarde van persoonsgegevens – inclusief gedragsgegevens – kan verhandelen? Dit kunnen zelfsturende systemen zijn onder beheer van de betreffende persoon of systemen

van commerciële partijen die dit in opdracht van het individu uitvoeren. Zogenaamde ‘banken’ voor persoonsgegevensuitwisseling. Als voorbode van hetgeen eraan komt, zagen we al in 1997 het Microsoft Passport, een eerste verschijning van een dergelijk systeem, bedoeld als *identiteitsbroker*. Facebook is een andere partij die identiteitmanagement met bijbehorende gegevens aanbiedt aan derden. Maar zoals met zoveel van deze toekomstvoorbidden, missen ze kritieke elementen. Om maar te beginnen met het feit dat ze nooit de uiteindelijke persoon als belangrijkste partij in dit systeem voor ogen hadden of hebben. Het staat dus te gebeuren dat een geheel nieuwe partij of, beter nog, een aantal partijen zullen verschijnen die in staat zijn om het vertrouwen van de consument als drijfveer van hun bestaan te beschouwen. Deze identiteitsbrokers zullen op basis van dat vertrouwen met voldoende slagkracht onderhandelen met gegevens-vragende organisaties over de waarde van de gegevens en de toestemming van de consument voor het gebruik van enkele of meerdere aspecten van persoonsgegevens. Het individu kan per transactie aangeven welke aspecten gedeeld mogen worden, onder welke voorwaarden en voor welke prijs, korting, of service. Wellicht in de vorm van een *smart contract*. Denk bijvoorbeeld aan het verstrekken van een leveringsadres voor een webshop-aankoop met een smart contract waardoor het adres na levering gedelete moet worden. Of aan het

aanmaken van een account op social media met een vergoeding voor elke keer dat jouw gegevens meegenomen worden in een advertentieverkoop. En niet te vergeten, de consument zal in staat zijn om middels het brokersysteem de toestemming (*Consent*) weer in te trekken, de gegevens over te dragen aan een derde partij (*Data Portability*) of eventueel te vragen om alle gerelateerde informatie, gegevens en afgeleiden daarvan, te vernietigen (*Right to Erasure*). Voor (commerciële) afnemers heeft een dergelijk marktsysteem ook enorme voordelen.

“Toch niet weer een privacydiscussie rondom Big Data.”

De gegevens worden bijgehouden door het individu zelf en dit zorgt voor een hogere kwaliteit en ‘versheid’ van de gegevens. Het brokersysteem maakt het ook vele malen eenvoudiger om te voldoen aan de wetgeving; Consent is duidelijk vastgelegd, Right to Erasure is makkelijker uit te voeren en Data Portability kan aan het brokersysteem uitbesteed worden. Daarnaast zullen de partijen die hiermee het eerst gaan werken een concurrentievoordeel halen uit het feit dat zij de waarde van persoonsgegevens juist inschatten en persoonsbescherming hoog in het vaandel hebben staan (*Privacy by Design*). In de toekomst zal de consument kunnen beslissen of organisaties hun Big Data met of zonder persoonsgegevens krijgen. Wie gaat hem of haar daarbij helpen?







2. DE STAAT VAN DATA

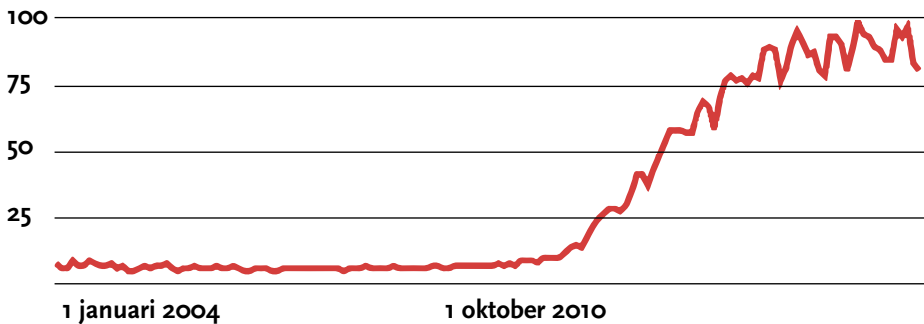
Er is een wildgroei aan nieuwe sensoren om informatie mee te verzamelen; opslagtechnologie wordt steeds kleiner en meer genetwerkt, computersystemen kunnen krachtiger en sneller analyses maken en slimme algoritmes gaan niet alleen de competitie aan met mensen, ze winnen deze ook vaak. Met de opkomst van datatechnologie die rekenkracht, opslag en algoritmische nauwkeurigheid maximaliseert, zijn mens-machinerelaties aan het verschuiven.

Tot voor kort stonden mensen pontificaal centraal in het kennis-productieproces. In het traditionele proces maken we kennis door op gestructureerde wijze informatie te verzamelen, deze op te slaan, data te categoriseren en een analyse te presenteren. Nieuwe technieken zorgen ervoor dat elk van deze stappen geautomatiseerd kan worden uitgevoerd door datasystemen: het verzamelen, registreren, opschonen, analyseren en presenteren van data worden parallel aan elkaar geautomatiseerd en de menselijke component in deze processen wordt stapje voor stapje gemarginaliseerd. We bouwen zo aan systemen die zelf een mate van intelligentie hebben en kunstmatig en onafhankelijk kennis kunnen produceren. De code die wordt gebruikt om datasets te analyseren ontwikkelt langzaam maar zeker het vermogen om zelf vragen te stellen, zelf patronen te herkennen en zelf oplossingen te bedenken. Datasystemen beginnen met ons mee te praten en de relatie tussen mens en data wordt hierdoor minder eenzijdig en meer interactief. Dit is ook noodzakelijk. Nu er dagelijks exabytes aan veelal ongestructureerde data worden geproduceerd (McAfee *et al.* 2012) is het onmogelijk en onwenselijk om de prominente rol van mensen in het kennisproductieproces te willen handhaven.

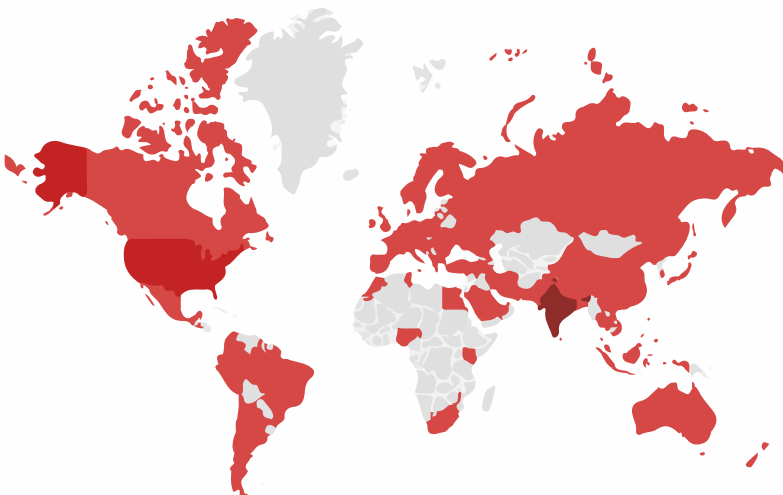
In dit hoofdstuk gaan we in op de vraag wat we bedoelen met de term Big Data, en hoe dit fenomeen zich de komende jaren zal ontwikkelen. We gaan in op Big Data als fenomeen dat een bepaald wereldbeeld veronderstelt, en ook op de technologische aspecten. Het Big Data-proces wordt hiervoor ontleend in vier stappen: het verzamelen van data, het opslaan van data, het analyseren van data, en het toepassen van data. Deze vier processtappen veranderen zelf niet of nauwelijks, maar de technologie die gebruikt wordt bij de individuele stappen is constant in verandering.

Wat is Big Data?

De manier waarop de wereld en haar complexe dynamiek in informatie kan worden gevat is altijd een uitdaging geweest. De snelheid waarmee de hoeveelheid opgeslagen informatie zich uitbreidt neemt alsmaar toe en dit fenomeen heeft de laatste jaren hernieuwde aandacht en ook hernieuwde labels gekregen. In 1997 schreven Cox en Ellsworth een artikel over datavisualisatie waarin ze stellen dat 'visualisatie een interessante uitdaging biedt aan computersystemen: datasets zijn over het algemeen erg groot, en belasten de capaciteiten van het geheugen, de lokale opslag en zelfs externe servers zwaar. We noemen dit het Big Data-probleem' (Cox & Ellsworth 1997). Dit was het eerste artikel dat Big Data als term gebruikte. In 2001 schreef Laney, een analist van de Meta Group, een onderzoekspaper over karakteristieke eigenschappen van Big Data en sindsdien heeft de term een vlucht genomen.



Figuur 2.1 Interesse in Big Data van 2004-2017, volgens Google Trends 2017.



Figuur 2.2 Interesse in Big Data op basis van regio, volgens Google Trends 2017.

Ook in Nederland zien we de term terug in honderden publicaties, conferenties en beleidsrapporten. Zo deed de WRR in 2016 onderzoek naar 'Big Data, privacy en veiligheid', om een normatief kader op te stellen voor het gebruik van Big Data door de overheid (Hirsch Ballin *et al.* 2016). In hetzelfde jaar liet het Ministerie van Economische Zaken een groep van privacy- en big-dataexperts advies geven om te komen tot verantwoord innoveren met Big Data (Expertgroep Big data en privacy 2016). Rathenau (2015) schreef twee rapporten over waarden als privacy, vertrouwen en solidariteit in een tijd van Big Data (Kool *et al.* 2015, Timmer *et al.* 2015). De Nationale Denktank (2014) schreef een rapport over Big Data en maatschappelijke toepassingen en ook het publiek-private COMMIT, diverse internationale overheidsorganen (o.a. The White House 2014, en Project Byte vanuit de EU), en een veelheid aan bedrijven en wetenschappers zijn met het thema bezig.

Hoewel er verwacht wordt dat de hype-term 'Big Data' op de lange termijn niet toekomstbestendig is, zijn de achterliggende ontwikkelingen dit wel degelijk. Het is daarom belangrijk dit concept onder de loep te nemen en aan te geven wat we bedoelen als we het over Big Data hebben. Volgens het artikel van Laney zijn 'drie V's' de hoofdkenmerken van Big Data (Laney 2001). De eerste V staat voor Volume en gaat over de grootte van databestanden en datasets. De tweede V staat voor Velociteit of snelheid waarmee data kan worden geanalyseerd en gebruikt. Ten slotte is er Variëteit of verscheidenheid aan data. De variëteit aan data die we produceren komt namelijk van verschillende actoren: we produceren zelf data, evenals onderzoekers, ambtenaren, professionals, bloggers en filmers. Daarnaast zijn er steeds meer 'dingen' die data verzamelen, zoals sensoren,



Figure 2.3 Big Data en de 3 V's.

camera's en huishoudelijke applicaties. Het gaat bij Big Data om grote hoeveelheden ongestructureerde informatie die van een veelheid aan bronnen komt. Om voordeel te halen uit Big Data wordt er op nieuwe manieren naar informatie gekeken. Traditioneel gebruiken we datasets waarin alle informatie vooraf gecategoriseerd wordt. Nu we toegang hebben tot snelle processoren, goedkope opslagtechnologie en krachtige netwerken kunnen we ook ongestructureerde, niet-gecategoriseerde informatie analyseren. Hierdoor kunnen we datasets koppelen die op eerste gezicht misschien irrelevant lijken, maar in tweede instantie een schat aan gecorreleerde informatie opleveren. We kunnen zo gericht hengelen in steeds grotere datasets.

In de literatuur zijn sindsdien een aantal nieuwe V's aan dit drietal toegevoegd, zoals *Value*, *Variability*, *Veracity* en *Virtual*. Belangrijk is dat het bij Big Data niet alleen draait om het volume, maar ook om het ecosysteem waarin data zich begeeft en de manier waarop we data koppelen en analyseren. Big Data kan vanuit wetenschappelijke perspectief gezien worden als een onderzoeksmethode – een manier om op kwantitatieve wijze data te verzamelen en om te zetten in informatie en kennis die ondersteund wordt door nieuwe datatechnologie en analysemogelijkheden.

Maar Big Data is niet alleen een technisch proces, betogen danah Boyd en Kate Crawford. Zij definiëren Big Data als “een cultureel, technologisch, en wetenschappelijk fenomeen” en stellen dat niet alleen technologie, maar ook een bepaalde mythologie een rol speelt: “het geloof dat grote datasets leiden tot een hogere intelligentievorm waarmee inzichten kunnen worden gegenereerd die voorheen onmogelijk waren en die een aura van waarheid, objectiviteit en accuraatheid hebben” (Boyd en Crawford 2012). Steve Lohr schreef dat we te maken hebben met een nieuwe filosofie, ‘data-isme’, en omschrijft dit als het geloof dat alles dat gemeten kan worden, gemeten hoort te worden. Lohr trekt de achterliggende assumptie, dat meten en weten een betrouwbare en transparante lens oplevert die subjectiviteit vermijdt, in twijfel (Lohr 2015).

Het geloof in de werking van data komt misschien wel het sterkst terug in een veel geciteerd artikel van *Wired Magazine* hoofdredacteur Chris Anderson. Hij stelde al in 2008 dat we met onze petabytes aan informatie het einde van theorievorming konden aankondigen: “Correlatie is genoeg, we kunnen stoppen met onze zoektocht naar modellen. We kunnen data analyseren zonder hypothesen op te stellen. We kunnen cijfers in de grootste computerclusters ter wereld gooien en statistische algoritmes patronen laten vinden waar de wetenschap dit niet kan” (Anderson 2008). Zowel de suggestie dat correlatie causatie kan vervangen als ook het idee dat data-analyse waarde vrij is, zijn veel bekritiseerd. Wel is het belangrijk in deze studie te verkennen wat voor epistemologische veranderingen Big Data-technologie brengt. Dat betekent dat we de vraag moeten stellen of Big Data de manier verandert waarop we de wereld (kunnen) kennen. Big

Data is al met al een containerbegrip. De WRR (Hirsch Ballin *et al.* 2016) heeft een aantal van de meest belangrijke kenmerken van Big Data uit de literatuur gehaald en presenteert deze in de onderstaande tabel. Big Data moeten we volgens de WRR zien als een samenspel van ontwikkelingen en niet zozeer als een vastomlijnd en definieerbaar gegeven. Er zijn niet veel praktijkvoorbeelden van toepassingen waarbij alle kenmerken uit de tabel tegelijkertijd aanwezig zijn. In de meeste gevallen gaat het om een combinatie van de kenmerken.

Kenmerken van Big Data (WRR, 2016)
<p>Data</p> <ul style="list-style-type: none"> • Omvang van de data: het gaat om grote hoeveelheden gegevens. • Structuur van de data: het kan gaan om gestructureerde of ongestructureerde gegevens of een combinatie van beide. • Variëteit van de data: het gaat om de combinatie van verschillende databronnen.
<p>Analyse</p> <ul style="list-style-type: none"> • Methode van analyse: de analyse is datagedreven, er wordt dus gezocht naar patronen in de data zonder vooraf opgestelde hypothesen. • Oriëntatie van de analyse: hoewel Big Data-analyses ook inzicht kunnen geven in het verleden (retrospectieve analyses), zijn het met name de analyses van het heden (<i>realtime-analyses</i>) en de toekomst (<i>predictive analyses / forecasting</i>) die de aandacht trekken.
<p>Gebruik</p> <ul style="list-style-type: none"> • Ontschotting van domeinen: data uit het ene domein worden gebruikt voor beslissingen in het andere domein. • <i>Actionable knowledge</i>: conclusies op geaggregeerd niveau kunnen worden toegepast voor beslissingen op groeps- of individueel niveau (persoon of object).

Big Data en technologie

Big Data is niet één aparte technologie; het is een combinatie van ontwikkelingen op het gebied van rekenkracht, netwerken, sensoren en datasets. In de volgende paragrafen bekijken we de technologie en de technologische trends die een rol spelen in het Big Data-proces. Deze kunnen worden ingedeeld volgens de volgende vier stappen die centraal staan in het Big Data-proces:



Verzameldrift

Toen in 1991 een aantal onderzoekers aan de Universiteit van Cambridge het beu waren om steeds maar naar de gedeelde koffiekant te lopen en die leeg aan te treffen, besloten ze actie te ondernemen. Ze gebruikten een camera en een Motorola 68000 computer, en maakten een netwerksensor die hen in staat stelde om de inhoud van de koffiekant te monitoren op hun beeldscherm. Niet veel later werd hun *Trojan Coffee Pot-server* een hit op het vroege internet en ontwaakte bij velen een droom. Kunnen we niet alles aansluiten op het internet? Inmiddels is deze droom verworpen tot een gigantische industrie, die wel wordt aangeduid als Internet of Things.

Een invloedrijke en vroege toekomstvisie hierop werd in 2001 opgesteld door het Institute for Prospective Technological Studies (IPTS) onder de noemer ISTAG-scenario's (Ducastel 2001). Het rapport, dat nog steeds actueel is, stelt een wereld voor waarin technologie alomtegenwoordig is. *Ambient intelligence* omringt ons, het is overal; niet alleen in onze huizen en kantoren, maar ook in onze auto's, in het openbaar vervoer, op straat, in het park, in het bos, in musea, winkels en restaurants. Netwerken van technologie communiceren continu met elkaar om ons als gebruikers op een zo 'natuurlijke' en onopvallend mogelijke manier van informatie en entertainment te voorzien. Om ervoor te zorgen dat we niet constant het gevoel hebben bezig te zijn met al die technologie is het de bedoeling dat technologie weggewerkt wordt: in deuren, ramen, kleding, stoeptegels, meubels, en zelfs in ons lichaam.

Een belangrijk kenmerk van ambient intelligence-technologie is dat het gepersonaliseerde diensten levert. Denk bijvoorbeeld aan schermen die voor het individu relevante route-informatie geven, of aan interactieve muren met advertenties waarbij een touchscreen ervoor zorgt dat men direct boodschappen kan bestellen. Bibi van den Berg (2010) schrijft hierover dat personalisatie vooral betekent dat de aangeboden hoeveelheid en het type informatie wordt aangepast aan de gebruiker. Het idee is dat men geen overbodige informatie of diensten krijgt aangeboden. Mannen hebben er geen baat bij te worden verwezen naar damestoiletten, maar willen misschien wel weten wat de temperatuur buiten is. Wie tijd over heeft zal veel infotainment ontvangen, terwijl men op haastige momenten behoed kan worden voor onderbrekingen. De visie achter de ontwikkeling van ambient intelligence streeft naar betrouwbare en veilige hardware die onopvallend is, natuurlijk aanvoelt en voorzien is van naadloos schakelende communicatie-infrastructuur. Volgens deze ontwerpprincipes bouwen we aan een uitgebreid internet of things dat als een zenuwstelsel de wereld monitort. Inmiddels zijn er verschillende voorspellingen die aangeven dat het aantal verbonden apparaten in de komende tien jaar zal uitbreiden tot 150 miljard apparaten met een marktwaarde van meer dan 10 biljoen euro (Chui *et al.* 2010, Evans 2011). Deze apparaten kunnen op afstand bestuurd worden, en vervangen – vaak op meer efficiënte, effectieve en goedkope wijze

– wat voorheen door mensen werd gedaan. De sensoren van de miljoenen apparaten beginnen menselijke zintuigen na te bootsen en geven door wat ze zien, horen, ruiken, voelen en meten. De voelsprietten van het Internet of Things zijn in staat mensen te identificeren, categoriseren, gedrag te analyseren, gezichten te herkennen, en hierbij ook emoties te herkennen of te liplezen (Janssen *et al.* 2015). Google's LipNet onderzoeksgroep uit Oxford voedde bijvoorbeeld haar systeem met duizenden uren beeldmateriaal van de BBC wat leidde tot een systeem dat 1.78 maal beter kan liplezen dan menselijke liplezers. Dit kan de komende jaren worden ingezet om dove mensen te helpen liplezen, om films te ondertitelen, om persoonlijke digitale assistenten fluisterend te besturen, of voor surveillance middels het bestuderen van camerabeelden.

Dit soort slimme sensoren veranderen de manier waarop we interacteren met de wereld. We praten tegen onze televisies en telefoons en ze luisteren ook nog eens naar wat we zeggen dankzij geluidsherkenningssensoren die in de cloud worden bijgestuurd om taalfouten te begrijpen en te corrigeren. Apparaten in de zorg sturen data zelfstandig naar artsen, en *fitbits* en andere apparaatjes monitoren voortdurend onze gezondheid. Op een bepaalde manier zijn we al cyborgs, of samensmeltingen van mens en machine, aangezien we gedeeltelijk offline en gedeeltelijk online leven en ons eigen lichaam, het zenuwstelsel en het brein worden ondersteund door technologie zoals onze mobieltjes en computers. Zoals het Rathenau Instituut en het SCP omschrijven zullen komende jaren naar verwachting nano-,bio-, informatie-, en cognitietechnologie verder convergeren waardoor biologie steeds meer technologie wordt en technologie steeds meer biologie (van den Broek *et al.* 2016). Zo bouwen we niet alleen aan een Internet of Things, maar gaan we verder en bouwen aan een intelligent Internet of Everything. En deze intelligentie wordt massaal opgeslagen als data.

Opslaan – denken als een database

Hoewel we niet veel merken van alle databases om ons heen, spelen ze al lange tijd een bepalende rol in onze informatiesamenleving. De manier waarop een database jou kent kan bepalend zijn voor de dingen die je doet en doelen die je hebt. Mag je het vliegtuig in? Als de database jouw naam herkent dan wel. Mag je stemmen? Dat bepaalt een database van de overheid. Elke keer als je een formulier ziet met vakjes waarop je iets moet invullen, communiceer je met een database. Iedere keer dat je een website gebruikt interacter je met een database, of misschien wel met een hele rits databases. Bol.com, Facebook, Twitter; het zijn allemaal databases. Iedere keer als je een 'computer says no'-scenario meemaakt, ligt er een database aan ten grondslag.

De databases van de toekomst worden nu op universiteiten voorgedookt. Zo presenteerden wetenschappers van de TU Delft in 2016 in *Nature* de kleinste harde schijf ooit, waarbij elke bit bestaat uit slechts één chlooratoom (Kalff *et al.* 2016). Deze harde

schijf heeft een opslagcapaciteit van 80 terabyte per vierkante centimeter, hoewel ze op dit moment alleen nog in een vacuümomgeving, gekoeld tot -196 graden Celsius, kan functioneren. “In theorie zouden we met deze capaciteit alle boeken die de mensheid ooit heeft geschreven kunnen bewaren op één postzegel”, stelt Sander Otte, leider van het onderzoek.

Ook wordt geëxperimenteerd met DNA als datadrager. Britse wetenschappers sloegen alle 154 sonnetten van Shakespeare, een audiobestand met de speech ‘I have a Dream’ van Martin Luther King en het artikel van Francis en Crick over de structuur van DNA, succesvol op middels DNA en konden deze vervolgens nagenoeg perfect uitlezen (Goldman *et al.* 2013). DNA kent een aantal voordelen. Het heeft ongelooflijke datadichtheid en kan tot 2.2 petabytes per gram opslaan. Daarnaast is het ontzettend duurzaam materiaal wat de data zeer goed beschermt tegen hitte en vochtigheid. Harvard onderzoeker George Church stelt dat je DNA “kunt droppen waar je wilt, in een woestijn of je tuin, en 400.000 jaar later zal het er nog net zo zijn” (Ibid.). Het is zodoende heel wat stabielere dan de onduurzame opslagtechnieken die we nu gebruiken. Maar ook deze methode is momenteel niet bruikbaar vanwege de kosten en klinische omstandigheden die nodig zijn om ermee te werken.

Hiernaast wordt geëxperimenteerd met helium gevulde harde schijven, nieuwe magneettechnologie, hologrammen waarbij data in talloze holografische lagen opgeslagen kunnen worden en zelfs met quantumopslag waarbij data zich tussen een staat van zijn en niet-zijn begeven.

De integratie van data

Hoewel vermogen en afmeting belangrijke eigenschappen zijn in de discussie over data-opslag, is er nog een ander essentieel aspect dat de mogelijkheden van Big Data vergroot: communicatie tussen data en databases. Als we alle kennis ter wereld kunnen opslaan op het formaat van een postzegel, maar deze niet snel, gemakkelijk en betrouwbaar kunnen overzetten naar een ander systeem, dan blijven de toepassingen nog steeds beperkt.

Om te speculeren over data-integratie die gaat komen, moeten we een kort uitstapje maken naar de jaren '70 toen de databases waarmee wij nu opgroeien werden uitgevonden. Dit type werkte volgens het relationele model, ofwel SQL, en dit ging gepaard met veel ijzerwaar: grote rekken met draaiende banden, analoge systemen, diskettes, silo's, grote ventilatoren, klinische ruimtes. De meeste content-managementsystemen op het web worden nog in de SQL-taal aangestuurd, terwijl het oorspronkelijk een taal is om tape-drives aan te sturen. Berichten hadden in deze databases een gefixeerde lengte, zoals SMS dat heeft, en net als SMS'jes hebben deze databases een strikte beperking: “als het niet in de database past, dan bestaat het niet”.

De opkomst van het internet bracht een revolutie. Met de opkomst van het wereldwijde web van Tim Berners-Lee kreeg de database een ander toekomstperspectief; er ontstond een netwerk van SQL-databases die informatie konden uitwisselen en zo betere samenwerking tussen organisaties mogelijk maakte. Maar de databases hebben eigenlijk nooit goed met elkaar leren omgaan. Ze konden (en kunnen nog altijd) niet goed met elkaar communiceren, waren niet gestandaardiseerd; de thuis-pc en de server op het werk werkten niet samen.

Omdat het internet zoveel databases met elkaar verbond ontstond al snel een nieuwe behoefte om alle databases te kunnen doorzoeken en een naald in een hooiberg te kunnen opsporen. SQL-systemen kunnen prima binnen één silo informatie achterhalen, maar niet binnen een miljoen silo's. Nieuwe infrastructures, met namen als NoSQL, Hadoop, of MapReduce, werden ontwikkeld. Dit zijn typische Big Data-systemen. Ze kennen een simpel ontwerp, zijn minder gestructureerd, zijn gericht op data-integratie en zijn in staat op flexibele wijze veel meer data snel te doorzoeken en rangschikken. Niettemin zijn er nog steeds veel problemen met het communiceren tussen systemen.

Er zijn verschillende modellen om netwerken en databases soepel aan elkaar te koppelen. De meest gangbare is om database A via een netwerk aan database B te koppelen en transacties tussen de twee te faciliteren. Dit is de meest logische manier om data te koppelen, maar ook een manier die intrinsieke problemen oplevert. Ten eerste is het systeem instabiel, omdat beide partijen verschillende software kunnen gebruiken en een klein foutje catastrofaal kan zijn voor de organisaties die met deze databases werken. De Mars Climate Orbiter is in 1999 verloren gegaan omdat Lockheed en NASA verschillende systemen hadden en het ene team met centimeters werkte en het andere met inches. Op afstand kun je niet met de broncode van de andere organisatie meekijken en zo is er niet altijd zekerheid over de data die je met elkaar deelt. Nu werkt men steeds meer met standaarden, maar je hoeft maar naar een standaardenregister te kijken om erachter te komen dat de ideale standaard niet gevonden kan worden. Opschalen is in dit systeem ook een probleem. Als één partij overgaat op nieuwe software, zal de andere partij ook mee moeten, en als er meerdere contacten zijn moet er in een heel netwerk aan systemen worden geïnvesteerd. Dit remt samenwerking en uitbreiding van het netwerk en dit is geen abstract hypothetisch probleem – het speelt bij bankieren, de zorg, logistiek en de overheid. Onze datasystemen zijn voorlopig nog een puinhoop.

Een alternatief dat in de startblokken staat is *blockchaintechnologie*. Dit is een gedistribueerde database die data opslaat en bijhoudt door middel van een consensus-systeem. Er is geen centrale partij of derde partij nodig om het vertrouwen in elkaar te garanderen en er hoeft dus ook geen commissie voor derde partijen te worden betaald. Met een blockchain wordt data gehard tegen manipulatie en vervalsing, doordat de consensus tussen alle 'nodes' garandeert dat gegevens juist zijn.

In theorie kan blockchaintechnologie een aantal van de problemen van onze huidige databases en netwerken oplossen. Het zorgt ervoor dat we binnen één systeem opereren, geen problemen krijgen met interoperabiliteit (communicatie tussen systemen), geen derde partij nodig hebben om data te valideren, en meer veilige en betrouwbare data hebben. Of blockchaintechnologie de komende jaren deze beloftes waarmaakt, valt nog te bezien.

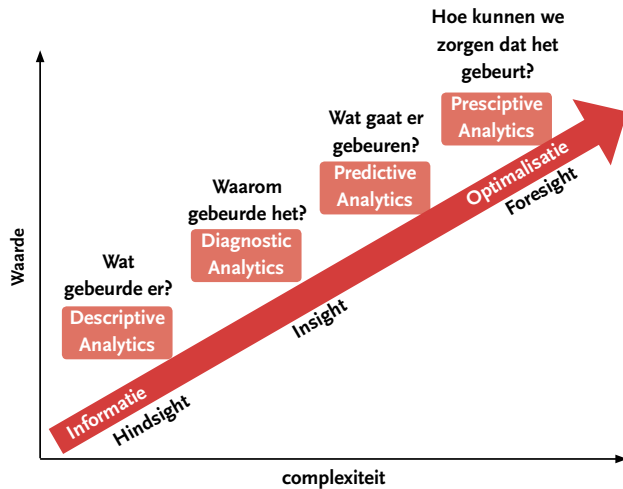
Analyseren – data als grondstof voor intelligentie

Met sensoren en opslagtechnologie kun je data verzamelen en bewaren, maar zonder een goede methode om deze data om te zetten in kennis creëer je geen meerwaarde. Bij Big Data zien we dat algoritmes¹ – gedigitaliseerde sets met regels die stap voor stap uiteenzetten welke handelingen moeten worden verricht om tot een doel te komen – een belangrijke rol spelen in de analyse van data en de operationalisatie van datasystemen.

Een vorm van geavanceerde algoritmes zien we bij de toepassing van machine learning, waarbij computers geen duidelijke codes meekrijgen, maar ‘getraind’ worden om zelf te analyseren en leren. Als je wilt dat een neurale netwerk leert wat een fiets is, dan geef je geen instructies om te letten op tweewielers, een kettingkast, of een fietsbel, maar je voedt het systeem duizenden plaatjes van fietsen. Als het daarna nog niet weet wat een fiets is, verander je niet de code, maar voer je nog meer plaatjes van fietsen toe. Naast plaatjes van fietsen wordt deze typische Big Data-technologie steeds meer ingezet om gezichten te herkennen, teksten te vertalen, om ongelukken te voorkomen bij zelflerende auto’s en zoekresultaten te vinden. Met deze techniek (*recurrent neural network language model* of RNNLM), werkt Google samen met de universiteiten van Stanford en Massachusetts aan systemen die visuele kunst kunnen produceren – zoals in dit boek te zien is – en ook woord voor woord zinnen kunnen bouwen. Zo krijgen we geautomatiseerde teksten die nog het meest lijken op poëzie (Bowman *et al.* 2016):

*there is no one else in the world.
there is no one else in sight.
they were the only ones who mattered.
they were the only ones left.
he had to be with me. she had to be with him.
i had to do this. i wanted to kill him.
i started to cry.
i turned to him.*

1 Vernoemd naar de Perzische wiskundige Al-Chwarizmi.



Figuur 2.4 Verschillende soorten data-analytics.

Verrassend aan dit soort technieken is dat ze helemaal niet vooruitstrevend zijn. Het zijn over het algemeen heel oude technieken die decennia geleden zijn bedacht. Het eerste generiek lerende algoritme dat werd ingezet voor deep learning werd in 1965 gepubliceerd door Ivakhnenko en Lapa (1966). Veel van de deep learning-architecturen zoals RNNLM, deep neural networks, en deep belief networks zijn jaren geleden bedacht, maar zijn recentelijk aan een opmars begonnen en hebben aanzienlijk succes met automatische spraak- en beeldherkenning. De succesfactor hierbij is Big Data. Doordat moderne computatiekracht zoveel sneller en krachtiger is kunnen de algoritmes en gelaagde neurale netwerken beter worden gevoed met data en beter worden toegepast.

Door de inzet van zelflerende systemen zijn er minder menselijke codeurs nodig en hoeft code niet constant herscheven te worden. Zoals Jeroen van den Hoven schrijft in zijn visiestuk (p. 43) hebben deze systemen het nadeel dat ze intransparant zijn. Je weet niet zeker of het systeem leert wat een fiets is, of dat alle foto's van fietsen toevallig buiten zijn gemaakt zodat het systeem een bepaalde omgeving of kleurenpatroon leert herkennen. Jason Tanz van *Wired Magazine* schrijft dat programmeurs hierdoor een nieuwe status verwerven als een soort hondentrainers, “en zoals iedere hondentrainer je kan vertellen, betekent dit dat je hele mysterieuze relaties aangaat” (2017). Een gevolg van de opkomst van intelligente datasystemen is dat de interactie tussen mensen en deze nieuwe intelligentievormen ruis en onbegrip kunnen opleveren. Dit neemt met de combinatie van Big Data en KI-functies zoals *machine learning* toe.

Datagebruik – applicaties

Recent is de KI-ontwikkeling in een stroomversnelling geraakt, aangezien de belangrijkste technologiebedrijven ter wereld hierin miljoeneninvesteringen hebben gedaan.

Er worden allerlei applicaties ontwikkeld die gebruikers op basis van data-analyses kunnen adviseren en die zelfs als persoonlijke assistent kunnen fungeren. In het consumentendomein wordt de opkomst van intelligente persoonlijke assistenten (IPA's) zoals Siri (Apple), Cortana (Microsoft), Echo (Amazon) en Google Now gezien als een van de snelst groeiende webdiensten (Hauswald et al. 2015). Dit leidt tot talloze nieuwe KI-platforms die een grote impact kunnen hebben op onze leefwereld.

Het Indiase MogIA voorspelde bijvoorbeeld als een van de weinigen de juiste uitkomst van de Amerikaanse presidentsverkiezing van 2016. Microsoft spraakherkenning scoorde beter dan mensen in het uitschrijven van gesproken dialogen. Een KI-systeem van de Future University Hakodate schreef samen met onderzoekers het verhaal 'The Day a Computer Writes a Novel' met als slotzin "The computer, placing priority on the pursuit of its own joy, stopped working for humans." Het korte verhaal werd door de jury van een literaire competitie in Japan in de eerste ronde verkozen als een van de winnaars en liet zo veel menselijke auteurs achter zich. KI en menselijke intelligentie zijn steeds lastiger te onderscheiden.

Vergelijk bijvoorbeeld eens de volgende twee teksten. Een van de teksten is geschreven door Emma AI en is speciaal voor deze publicatie opgesteld, de andere komt uit een rapport over de toekomst van KI door het Witte Huis (National Science and Technology Council 2016). Emma AI wordt door ontwerper Shaunak Khire omschreven als een 'crowd-sourced, supervised convolutional neural network that is designed to pass the Goertzel and Nills tests' (persoonlijke communicatie) en kreeg voor deze opdracht honderd artikelen en rapporten over Big Data en Kunstmatige Intelligentie ingevoerd, die geselecteerd waren op hun hoge citatiescore.

A: Many areas of public policy, from education and the economic safety net, to defense, environmental preservation, and criminal justice, will see new opportunities and new challenges driven by the continued progress of AI. Government must continue to build its capacity to understand and adapt to these changes. As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable.

B: Private and public institutions need to be encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit society. All sectors face the challenge of how to diversify the AI workforce. The lack of gender and racial diversity is worrying and will allow biases to seep through in the outputs of neural networks.

Is het duidelijk wie welke tekst schreef? De volledige bijdrage van Emma AI is te vinden op p. 48 en zo kan men ook achterhalen welk stuk tekst niet van Emma is.

De mens als onvoltooid project

Ik ben geen fan van het woord privacy. Het verdoezelt zaken die heel wezenlijk zijn, zowel nu als in de toekomst. Laten we het woord eens even weggooien. Waar het mij om te doen is, is dat ik jou kan schaden als ik informatie over je heb. Als ik genoeg van je weet kan ik geld van je rekening halen, ik kan weten waar je bent, of ik kan je onrechtmatig een baan weigeren vanwege een depressie die je ooit had. Privacy gebruiken we als label om dit soort belangen bij elkaar te pakken, maar een label kan ook verbergen wat erachter schuilt. Als iedereen doorheeft dat het bij privacy gaat om de mogelijkheid tot manipulatie, discriminatie, exploitatie van persoonsgegevens, en andere financiële, fysieke of aan reputatie gerelateerde schade, dan hebben we een andere discussie.

Morele autonomie

Naast mogelijke schade is er echter nog een aspect van privacy waarover we in gesprek moeten, namelijk het recht om jezelf te zijn of morele autonomie. Een fatsoenlijke samenleving laat mensen voor een belangrijk deel het verhaal over zichzelf vertellen. Als ik jou respecteer probeer ik te begrijpen wie jij wilt zijn en hoe jij je presenteert. Soms klopt iemands verhaal natuurlijk niet, maar binnen een zekere bandbreedte mogen mensen zich

presenteren zoals ze willen en zo hun eigen biografie schrijven. Big Data verleidt ons om te denken dat we alles over een persoon kunnen weten: iemands klikgedrag, locatiegeschiedenis, hersenscans, het genoom. Alles bij elkaar opgeteld geeft dat ons gemakkelijk het idee dat ik kan claimen te weten wie jij bent en wat goed is voor jou. Dat gaat voorbij aan de grens van iemands autonomie. Ik ben om die reden voorstander van kennisbescheidenheid. Voor elk mens is er een subjectief perspectief, en net zo goed als je niet met de stem van een ander kunt spreken, kan een ander niet precies in jouw schoenen staan en weten wat het is om jou te zijn. Elke karakterisering op basis van Big Data is een karakterisering van buitenaf en gaat voorbij aan een onreducerbaar subjectief perspectief. Dat is wat mensen bedoelen als ze zeggen: “mensen weten niks van mij.” Zelfs met terabytes aan data moeten we bescheiden zijn. Misschien heb je gisteren wel besloten om het helemaal over een andere boeg te gooien. Of ben je alles wat je bent met tegenzin. Mensen zijn onvoltooiden projecten, en ze moeten ook als zodanig opgevat worden.

Vals dilemma

We hebben de belangen van privacy redelijk netjes in wetten vastgelegd en

vormgegeven. We willen naast privacy echter ook onze banen en economische groei behouden, innovatie bevorderen, wetenschappelijk onderzoek doen, en terroristische aanslagen verijdelen. Dat kan op gespannen voet met elkaar staan. Denk aan de discussie over slimme elektrische meters, of over het elektronische patiëntendossier. Veel mensen capituleren bij deze conflicterende waarden en denken dat het gaat om een keuze tussen privacy of geen privacy, maar dat is een vals dilemma. We kunnen een derde weg inslaan door dit vraagstuk als innovatieopdracht te zien. In Duitsland zie je dit in het debat over ecologie en economie. Ze bewijzen dat het niet òf milieu òf economie is en zijn daarmee kampioen *renewable energy* en *cleantech* geworden. Zo kan dat voor privacy ook. Er is veel winst te halen als we in Nederland *privacy by design*, en *privacy respecting* of *privacy enhancing technologies* voorstaan. Dit geldt in het bijzonder voor Big Data-technologie. We gaan hiermee onmiskenbaar veel leren over hoe de samenleving werkt, wat des te nijpender is nu die samenleving zo complex is en nieuwe kwetsbaarheden kent. Je moet dus die Big Data gebruiken, maar we willen daarbij niet onze grondrechten opgeven.

Ter illustratie

Het ontwikkelen van datatechnologie volgens deze ontwerpprincipes gebeurt al. Stel dat je wilt weten hoeveel mensen er in een bepaalde ruimte zijn. Je kunt dan gebruik maken van camera's die automatisch de personen *blurren*, zodat je wel het aantal mensen kunt tellen, maar niet kunt achterhalen wie het zijn (*coarse graining*). Ook zie ik veel mogelijkheden voor

toepassingen die gebruik maken van *blockchain*. Daarnaast zijn er veel ontwikkelingen zoals *price-targeting* en *flexible pricing* waarbij er nog veel ethische vragen zullen opduiken. Bij de supermarkt verdwijnen de prijzen van de producten en weet slechts de scanner nog wat een brood kost. Als die scanner precies weet wie jij bent en wat je bereid bent voor een brood te betalen, dan kan die kennis tégen je gebruikt worden.

Big Data verleidt ons om te denken dat we alles over een persoon kunnen weten.

Online reisbureaus die uit mijn tweede bezoek weten dat ik naar Venetië wil vliegen kunnen de prijs verhogen. Dat zijn dingen die het grote publiek zich steeds meer realiseert en waar privacy in het geding is.

Hobbels op de weg

Een van de grote hobbels is dat de publieke sector relatief laat is begonnen om deze uitdaging op te pakken. Veel is ons door het bedrijfsleven, vooral internationale bedrijven, al uit handen genomen. Ik denk daarbij aan de vier of vijf spelers die zich de afgelopen tien jaar vol hebben laten lopen met onze data. We hebben die data eigenlijk vrij naïef weggegeven. Deze spelers maken zich op basis van onze data meester van superieure analysemodellen en technieken waarmee ze gedrag kunnen voorspellen en een nog dominantere rol in ons leven kunnen spelen. Ik maak me daar absoluut zorgen over en zie hier een rol voor de overheid weggelegd.

Misschien moet er gedacht worden langs de lijnen van het grotere publieke belang, zoals er eind negentiende eeuw bijvoorbeeld

publieke bibliotheken of openbare nieuwsvoorzieningen werden opgetuigd. Deze informatie-architectuur zorgde ervoor dat het grote publiek toegang had tot belangrijke informatie en dat ze beschermd werd tegen al te hebbelijke derde partijen. Overigens niets ten nadele van commercie en zakelijke belangen, maar we moeten wel voorkomen dat we alleen nog maar dansen naar de pijpen van winstmaximalisatie als het om de inrichting van de samenleving en de grondslagen van menselijke verstandhouding gaat. Als mensen de kans krijgen krijgt bij het meel te doen en er geen organisaties en instanties zijn die dat verbieden of er toezicht op houden, dan gebeurt dat morgen weer.

Uitlegbare systemen

In nieuwe privacywetgeving staat dat de werking van algoritmes en profielen uitlegbaar moet zijn, maar dit is technisch een enorme uitdaging. Je kunt die uitlegbaarheid wel in de wet opnemen, maar wat betekent het in de praktijk? Je zult vanuit de complexe systemen en kunstmatige intelligentie heel veel vertaalslagen moeten maken voordat de gewone consument de handelingen van een datasysteem kan begrijpen. Als het ten principale al kan. Er zijn immers nu al algoritmes waarbij de uitkomst voor eenieder ongewis is. Een voorbeeld uit recent onderzoek: een zelflerend systeem dat werkt op basis van neurale netwerken krijgt honderden plaatjes waarbij ofwel een husky ofwel een wolf is afgebeeld. Na verloop van tijd kent het systeem het verschil en categoriseert nieuwe plaatjes altijd juist. De wetenschapper denkt dat het systeem heeft geleerd wat een husky is, maar uiteindelijk blijkt dat

alle plaatjes van de huskies in een groene tuin zijn genomen, en alle plaatjes van wolven in de sneeuw. Het systeem heeft dus niks geleerd over wat huskies en wolven zijn, maar kijkt blijkbaar alleen naar de aanwezigheid van sneeuw om te bepalen of het beest een wolf is of niet.

We weten niet altijd hoe een systeem kijkt of wat het leert als het leert. Maar bepaalde processen zijn te belangrijk om over te laten aan *black-box-technologie*. Dat betekent dat we in die gevallen het gebruik van neurale netwerken moeten verbieden of sterk reguleren. Zoals er bijvoorbeeld een classificatiesysteem bestaat voor zeeschepen, waarin we vaststellen of een schip wel of niet geschikt is voor het IJsselmeer, de Baltische zee, of de Atlantische Oceaan, zo hebben we iets soortgelijks nodig voor data-technologie en digitale technologie in het algemeen. Bepaalde algoritmes zijn prima geschikt voor industriële processen, maar niet om HR-dossiers mee te beoordelen. We moeten misschien niet alleen zoeken naar universele regels, maar vaststellen dat er sectoren en situaties zijn waarin je bepaalde algoritmes of KI-technieken wel of niet mag gebruiken. Vergeet niet dat de meeste sectoren door schade en schande wijs zijn geworden. Denk aan de vliegwereld of de voedsel- en farmaceutische industrie. Er zijn in het verleden verschrikkelijke ongelukken gebeurd, maar door instituties, wetgeving, monitoring en inspectie is er in de samenleving vertrouwen opgebouwd. In de Big Data-sector gaan ontwikkelingen erg snel en is de regie ons min of meer uit handen genomen. Het is daarom tijd dat de overheid hier ingrijpt, en snel een beetje.

Even googelen wat de toekomst brengt...

We beschikken allen over informatie en kennis over de wereld om ons heen. Deze vormt de basis van onze handelingen en meningen, van wat we op internet met elkaar delen of 'liken'. Een slimme optelsom van deze geïnformeerde meningen en kennis is nauwkeuriger dan de beste inschattingen van experts. Dit is de *wisdom of crowds*: samen weten we meer.

We maken dagelijks gebruik van deze collectieve intelligentie als we iets opzoeken op internet. Google's zoekmachine kijkt daarvoor naar kruisverwijzingen die internetpagina's naar elkaar maken en die de data ordenen en structureren. Verder weegt Google de reputatie van websites en bronnen mee in de rangorde van de zoekresultaten.

Daarbij moet Google corrigeren voor reclamemakers die proberen websites hoger te laten scoren. Minstens zo'n grote uitdaging vormen het kudde- en kopieergedrag op het web: onjuiste informatie en leugens verspreiden zich razendsnel. Met de huidige groei van sociale media wordt het steeds lastiger om de juiste informatie te filteren. We praten elkaar steeds meer na waardoor diversiteit en onafhankelijkheid, essentiële voorwaarden voor de wisdom of crowds, verloren gaan.

Prediction markets: geld op het spel

In de praktijk blijkt dat de voorspellende waarde van schattingen en meningen verbeterd kan worden door gebruik te maken van zogeheten *prediction markets*. Deelnemers worden hierbij gestimuleerd om een correcte voorspelling te doen omdat ze dan een financiële beloning ontvangen.

In dit geval worden er vragen voorgelegd over een toekomstige gebeurtenis waarbij gekozen moet worden uit enkele antwoorden of uitkomsten. Bijvoorbeeld: wie wordt de nieuwe president van Amerika: Trump

Onjuiste informatie en leugens verspreiden zich razendsnel.

of Clinton? Deze antwoorden worden als aandelen op een beurs verhandeld. De antwoorden op de gestelde vragen worden gegeven door de koers van de aandelen. Hoe hoger de koers van het aandeel Trump of Clinton, hoe waarschijnlijker het is dat die kandidaat de winnaar van de Amerikaanse presidentsverkiezingen wordt.

De gedachte achter prediction markets is dat alle informatie en kennis die aanwezig is onder de deelnemers verdisconteerd is in de koers van het aandeel. Iedere

'handelaar' streeft naar maximale winst. Zodra hij nieuwe informatie heeft op basis waarvan hij de kansen van Trump hoger inschat, zal hij proberen aandelen bij te kopen en zal de prijs stijgen. De koers is daarmee een slimme optelsom van alle op de beurs aanwezige informatie en kennis, een vorm van collectieve intelligentie dus. Op deze manier is het mogelijk om behoorlijk betrouwbare antwoorden te krijgen op vragen over toekomstige gebeurtenissen.

Onafhankelijkheid dankzij de blockchain

Toch bleek in 2016 bij de laatste Amerikaanse presidentsverkiezingen en de *Brexit* dat prediction markets, die het tot nu toe beter deden dan opiniepeilingen, er flink naast kunnen zitten. De verklaring hiervoor lijkt te liggen in de homogeniteit van de groep handelaren: het is telkens hetzelfde type mensen dat deelneemt. Daardoor kunnen deze markten last krijgen van tunnelvisie en *groupthink*: alle deelnemers beschikken over ongeveer dezelfde informatie en hebben de neiging het met elkaar eens te zijn. Daarbij ondervinden deze aandelenbeurzen last van strategisch gedrag van deelnemers die willen profiteren van koersschommelingen.

De laatste jaren worden er diverse pogingen gedaan om de werking van prediction markets te verbeteren door het veel gemakkelijker te maken om mee te doen en door slimmere manieren van belonen die niet aanzetten tot strategisch gedrag.

Bedrijfjes als Augur [www.augur.net], Gnosis [www.gnosis.pm] en HiveMind

[bitcoinhivemind.com] maken daarvoor gebruik van blockchainplatformen zoals Ethereum en BitCoin. Met blockchain-technologie kan informatie decentraal opgeslagen worden in vele kleine versleutelde blokjes. Er is geen centrale marktplaats meer nodig en deelnemers kunnen individueel worden beloond. Zo wordt het mogelijk om individuen onafhankelijke voorspellingen te laten doen over de uitkomst van gebeurtenissen op basis van hun eigen specifieke en lokale kennis.

Een zoekmachine voor de toekomst

Door de decentralisatie van de aandelenmarkt kan in principe iedereen, waar dan ook ter wereld, expert of leek, hoogopgeleid of laagopgeleid, meedoen. Op deze manier kan meer kennis en meer diversiteit worden aangeboord. Ook kan hierdoor de 'handel' op de markt levendiger worden, zodat er een betere prijsvorming plaatsvindt en dus een betere collectieve inschatting van de uitkomst – de toekomst. Het is zelfs denkbaar dat ook intelligente apparaten en machines – robots – in de toekomst inbreng leveren aan zo'n kennissysteem.

Met deze verbeteringen wordt het mogelijk om een kennissysteem te bouwen dat gebruik maakt van alle bij mensen en machines aanwezige collectieve intelligentie. Aan zo'n systeem zou je vragen kunnen stellen over de toekomst in plaats van het verleden. Even googelen wat het nieuwe jaar je gaat brengen...

The future of AI

In 1950, Turing expected that by the year 2000 humans would be able to speak of thinking machines. By 2017, the concept of Artificial Intelligence (AI) was gradually appropriated to describe all sorts of unrelated and generic computer programs such as game players, search engines and chatbots. Now there is again an implication that an Artificial General Intelligence system (AGI) is just a smarter species of chatbot. Some people are wondering whether we should welcome our new robot overlords, and in that context the White House, which houses the President of the United States or the most powerful person in the free world, recently commissioned a report on the future of AI from The National Science and Technology Council. There is no single definition of AI that is universally accepted. This diversity makes it difficult to clearly define between what constitutes AI and what does not. Another challenge in using machine learning is that it is typically not possible to extract or generate a straightforward explanation for why a particular trained model is effective. It is therefore necessary to create mathematical structure that characterizes a range of possible decision-making rules with adjustable parameters. In recent years, some of the most impressive advancements in

machine learning have been in the subfield of deep learning, also known as deep network learning.

It is however important to consider that data is not the sole requirement for building an optically equal system. It is empirically clear that math and not data is responsible for the way a neural network is powered. Among the many types of neural networks available the most popular are Long Short Term Memory (LSTM) networks, Convolutional Neural Networks and more recently Bayesian Neural Networks. Deepmind AlphaGo, which beat Lee-Sedol in early 2016 at Go, used a custom LSTM network. Watson, IBM's Jeopardy winning AI, uses Convolutional Neural Networks (CNN) for Natural Language Processing (NLP) and Unstructured Information Management Architecture (UIMA) to power interaction models.

AI has made significant inroads in industries like finance, transportation, cybersecurity and personal assistants for consumers. In transportation, a 2013 study from the Association for Unmanned Vehicle Systems International predicted that the commercial drone industry could generate more than \$82 billion for the U.S. economy. In finance, according to David Siegel, founder of \$50 billion hedge fund Two Sigma, investment

management doesn't require the most esoteric features of the human brain. Numerai, an AI startup that does not disclose performance uses a method similar to homomorphic encryption. Meanwhile algorithms are evolving into full blown vertically integrated AI's using neural networks at Sentient Technologies, which again refuses to disclose performance.

State and non-state actors have focused on the use of artificial intelligence to create robot armies in both the online and the offline world. In the offline world it is very likely that autonomous military drones are operationalized within the next 5 years for the purpose of ethnic cleansing by non state actors and protection of civilians through

targeted strikes by state actors and so forth. Another use of AI in cyberspace is to create autonomous cyber attacks using a combination of IOT devices that could be turned into botnets. An AI that is programmed by non-state actors can achieve an objective and destabilize a nation's critical infrastructure using LSTM networks, Bayesian networks and open source code. As a result it is highly unlikely that there will be international co-operation between countries stemming from the inherent distrust between international regulators.

Private and public institutions need to be encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit

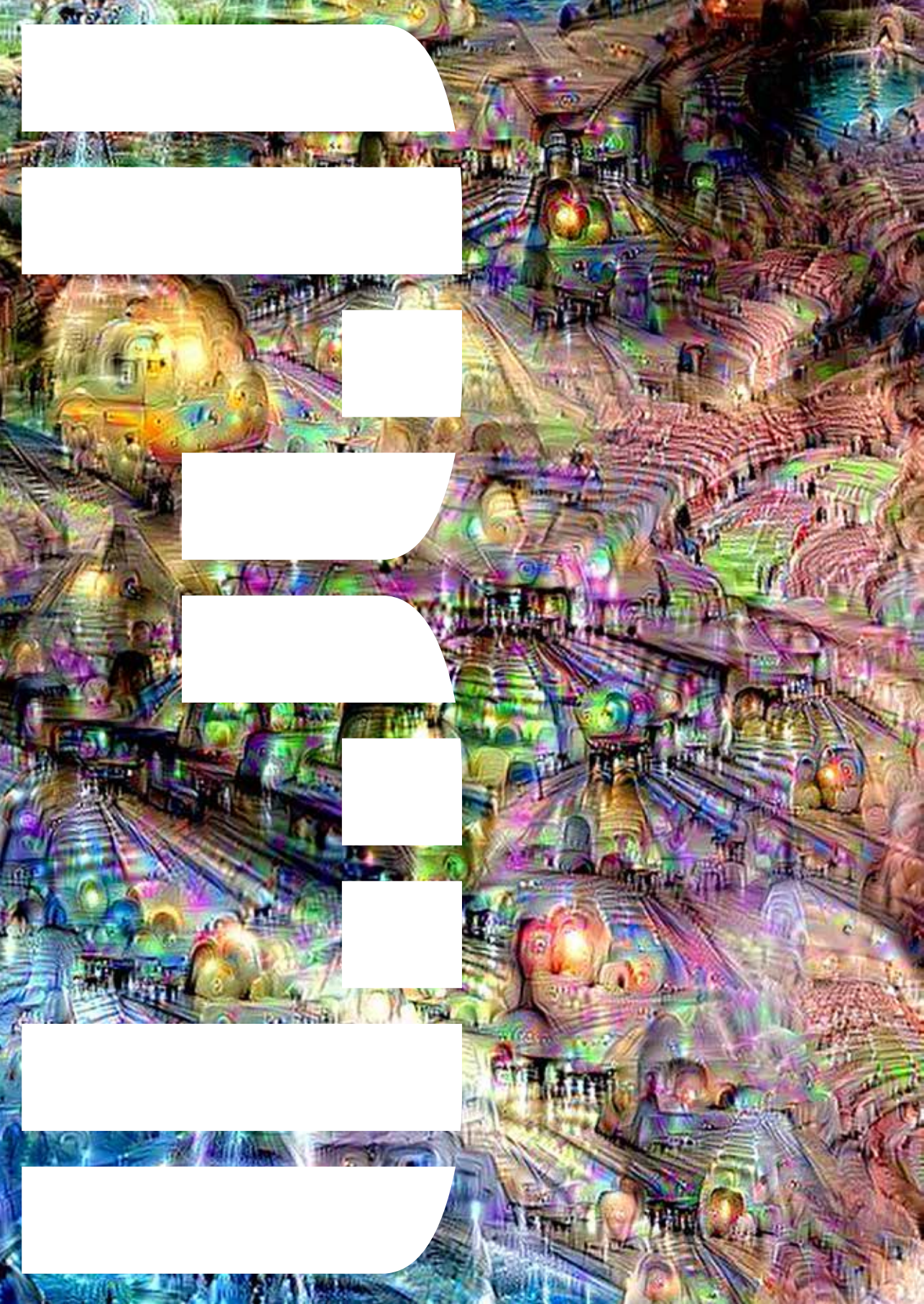
society. All sectors face the challenge of how to diversify the AI workforce. The lack of gender and racial diversity is worrying and will allow biases to seep through in the outputs of neural networks. AI does not need good data but rather a better way to find anomalies within a dataset. It is important that anyone using AI or machine learning is aware of this. Schools and universities should include ethics, and related topics in the context of AI. AI can be a major driver of economic growth, however there is also the threat that AI will replace

An AI that is programmed by non-state actors can achieve an objective and destabilize a nation's critical infrastructure using LSTM networks, Bayesian networks and open source code.

the workforce. Historically automation has threatened lower-wage jobs and with AI doing more and more knowledge intensive tasks it can destabilize white collar jobs as well. In the UK, research conducted by Oxford University and Deloitte, found that 1.3m administrative jobs across the public sector had the highest chance of being automated. Ultimately the economic effect of AI doesn't really have much to do with the development of a master AGI. The battle between good and evil ideas is as old as our species. What we need are intelligences with good ideas with a clear objective to defeat evil forces. However, the human race is fallible, and our own conception of 'good' needs continual improvement.







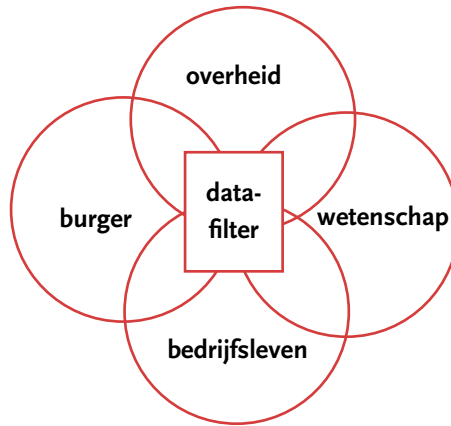
*The real danger is not that computers will begin to think like men,
but that men will begin to think like computers.*

Sydney Harris

3. WAT TELT? OVER DATA EN SOCIALE GEVOLGEN

De belangrijkste vraag voor de eenentwintigste eeuw, stelde futuroloog Alvin Toffler, omvat slechts enkele woorden: “Wat betekent het ‘mens’ te zijn?” (In Klaassen *et al.* 2009). Wat betekent het mens te zijn in een tijdperk van Big Data? De opmars van datatechnologie is niet alleen een technologisch proces, het heeft ook gevolgen voor de manier waarop we in de toekomst met elkaar omgaan, hoe we zullen handelen en wat voor mensen we zullen zijn. Big Data geeft zowel het bedrijfsleven, de wetenschap, de staat, als ook burgers mogelijkheden om elkaar *real-time* te monitoren, te analyseren en elkaars gedrag te voorspellen. Het reikt zo een nieuwe lens aan om elkaar te zien, om elkaar te categoriseren en om elkaar te begrijpen. Het helpt mee om nieuwe types mens te creëren, mensen die datagericht zijn en datagedreven te werk gaan. Burgers doen bijvoorbeeld aan citizen science door gegevens zelf te meten, delen, analyseren en toetsen. Data-analyse op basis van statistische correlatie pleit voor een nieuw type ambtenaar die een andere achtergrond en andere vaardigheden heeft dan de ambtenaar die we gewend zijn. In bedrijven zien we dat Chief Data Officers (CDO's), Chief Digital Strategists en zelfs KI-systemen het bestuursniveau bestieren. Bedrijven die zich oorspronkelijk niet met ICT bezigden herpositioneren zich rap als databedrijven.

Naast een nieuwe bron van inkomsten – data worden veelal vergeleken met goud of olie en steeds meer op deze manier behandeld en verhandeld – creëert het datagedreven leven ook nieuwe ‘sociale’ relaties. Niet alleen tussen personen onderling, maar ook tussen mensen en datasystemen. Onze relatie met overheden en met elkaar verschuift in de toekomst mogelijk naar een driehoeksverhouding waarin intelligente informatiesystemen een bemiddelende rol spelen. Zowel burgers, bedrijven als overheden zullen in hun relatie tot elkaar geadviseerd en gemedieerd worden door slimme systemen (zie schema 3.1). Het is belangrijk van deze systemen te weten hoe ze opereren, hoe ze leren en met welk doel ze tot stand zijn gekomen. De vraag wie temt en wie getemd wordt, is daarbij op zijn plaats.



Schema 3.1 Quadruple helix met centrale rol data.

Met de technologische ontwikkelingen uit het vorige hoofdstuk in gedachte gaan we in dit hoofdstuk verder in op de belangrijkste sociale gevolgen die Big Data op de samenleving zal hebben. Leidt de inzet van datagedreven technologie tot een ander begrip van het privacyconcept? Hoe kunnen we onze persoonsgegevens beschermen in deze tijd van *data-analytics*? Hoe geven we onze relaties met slimme systemen vorm?

Richards en King bieden in deze discussie een framework om de belangrijkste sociale implicaties van Big Data te onderzoeken (2013). Ze doen dit aan de hand van drie spanningsvelden. Ten eerste observeren ze dat Big Data bijzonder goed in staat is om personen te identificeren, maar dat het daarmee tegelijkertijd het individu kwetsbaar maakt. Dit noemen ze de identiteitsparadox. Ten tweede benoemen ze de transparantieparadox, die stelt dat we enerzijds steeds meer transparantie en inzicht krijgen, maar anderzijds te maken hebben met systemen die steeds minder transparant en begrijpelijk zijn. Het derde spanningsveld is de machtsparadox, die stelt dat Big Data de potentie heeft om de samenleving, het bedrijfsleven en de overheid te transformeren, maar dat dit ook kan leiden tot machtsverhoudingen die uit balans zijn. In dit hoofdstuk gaan we dieper in op deze spanningsvelden en worden onderwerpen als de toekomst van privacy, transparantie en vertrouwen in technologie behandeld.

Data en ik. Over Big Data en identiteit

Een belangrijk spanningsveld in de Big Data-discussie ligt tussen personalisering en privacy. Personalisering is het proces waarbij een systeem wordt afgestemd op het profiel en de context van een gebruiker. Om diensten te verlenen die zo persoonlijk mogelijk zijn is het belangrijk om rijke en accurate data over de gebruiker te verzamelen en te analyseren (Habegger *et al.* 2014). Deze informatie kan persoonskenmerken bevatten of als persoonlijk gezien worden en dit leidt ertoe dat een discussie over personalisering

al snel een discussie over privacy wordt. De komende paragrafen behandelen de identiteitsparadox en thema's zoals personalisering, profilering, datadiscriminatie en privacy.

Personalisering

Misschien kent u de sketch van cabaretier Hans Lebbis, waarin hij een perfect terrasje beschrijft. De scène wordt veel gebruikt in seminars over klantgericht handelen, maar symboliseert ook een Big Data-droom. De sketch gaat ongeveer als volgt. Je nadert een café, er staat fijne muziek op, de temperatuur is precies goed en er is nog één tafeltje vrij. Je gaat zitten, de serveerster komt op het moment dat je je jas uit hebt, en je vraagt haar om een glas rosé. Ze brengt een ijskoud glas waar het water vanaf parelt, precies zoals je het had willen hebben. Ze neemt de kaart mee en direct valt je blik op de salade niçoise, je favoriete salade. Niet veel later wordt de salade gebracht, met ansjovis erin, en je waant je dichtbij zee. Het ziet er goed uit, maar als de serveerster wegloopt denk je: “mayonaise! Ik wil graag mayonaise voor bij het ei.” De serveerster draait zich om en voordat je het weet presenteert ze een schaalje en zegt: “de mayo, uit eigen keuken.” Je doopt het ei erin en het is goddelijk; perfecte mayonaise. Je geniet en nadat je de salade op hebt wordt er netjes afgeruimd en een doekje over de tafel gehaald. Niet alleen een nat doekje, maar ook nog een droge zodat je niet – zoals laatst – met je arm in het water zit. “Wilt u verder nog iets?”, vraagt de serveerster. Je hebt weinig tijd, maar ze blijken je favoriete koffie te serveren. Bij het afrekenen zegt de serveerster €14,10 en je legt een briefje van €20,- neer. Je hoort jezelf zeggen “het is wel goed zo”, maar terwijl je dit doet realiseer je je dat 6 euro wel erg veel fooi is op zo'n bedrag. Je kijkt, de serveerster bloost een beetje, en je zegt: “nee, het is echt goed zo, het was perfect.”

“Waarom zijn dit soort terrasjes er nou niet?”, vraagt Lebbis tot slot en tot groot amusement van de zaal.

Bedrijven die met Big Data experimenteren, streven naar perfecte terrasjes, of beter gezegd, perfecte klantervaringen. Maar waar in het scenario van Lebbis een goede serveerster en wat toevalligheden leiden tot een perfecte ervaring, worden klantdata gezien als de technologische heilige graal om dit mogelijk te maken. Zodra een bedrijf toegang heeft tot een profiel waarin opgeslagen staat welke koffie, muziek, of temperatuur jouw voorkeur heeft kan het hiernaar handelen. Bestel je altijd mayonaise bij je ei, dan kan dit zonder het te vragen geregeld worden. Heb je de afgelopen keren nooit een maaltijd met vlees besteld, dan kunnen vleesgerechten minder opvallend worden aangeboden, of misschien zelfs van je persoonlijke menukaart verdwijnen. Heb je in het verleden

geklaagd over een natte tafel, dan kan men een extra droge doek over de tafel halen. Geniet je van extra koude rosé en verhoogt het schenken hiervan de kans dat je een fooi geeft? *Computer says yes!*

Door data uit voorgaande ervaringen te bundelen en te verrijken met informatie van sociale media, gedrag van disgenoten, of persoonlijke dossiers wordt het mogelijk om ervaringen te creëren die nauw aansluiten op de wens van de klant. Om de ervaring meer impact te laten hebben is dit soort technologie vaak verborgen en heb je niet door of en welke informatie precies verzameld en gebruikt is. De ervaring kan zo heel authentiek overkomen en interacties worden als toevallig, kloppend of perfect ervaren. Door klantgedrag te analyseren en profielen op te stellen kunnen organisaties uitproberen welke klantbenadering het beste werkt. Bovendien kunnen automatische sensoren door middel van bijvoorbeeld gezichtsherkenning steeds makkelijker achterhalen wie jij bent om zo je dossier te laden en te operationaliseren. Op terrasjes is dit misschien nog toekomstmuziek, maar online is het een bekend fenomeen.

“Big Data, aangenaam!”

Big Data brengt bedrijven steeds dichterbij de klant, en andersom. Warenhuizen als Amazon en Bol.com geven je op basis van je aankoopgeschiedenis suggesties voor producten. Talloze videodiensten schotelen je op basis van eerdere keuzes en jouw waarderingen nieuwe films en series voor. Online videodienst Netflix gaat een stap verder en creëert producten en content aan de hand van Big Data-analyses. “We kunnen kijken naar consumentengegevens en zien welke regisseur, welke filmster, en welk type film en shots populair zijn”, vertelt vicepresident Steve Swasey (In Lawler 2011). Zo bepaalt Netflix aan de hand van consumentenvoorkeuren ook wat voor trailer de kijker te zien krijgt. Als je eerder veel Kevin Spacey films hebt gekeken zie je Spacey terug in de trailer, als je van romantiek houdt wordt op romantische shots ingezoomd en ben je fan van de regisseur, dan wordt zijn naam benadrukt. Big Data leert je, vaak ongevraagd, kennen. Met een klein aantal Facebook-‘likes’ kan men al informatie achterhalen over iemands seksuele voorkeur, religieuze en politieke oriëntaties, persoonlijke karaktereigenschappen, gebruik van verslavende middelen én of hij of zij is opgegroeid in een gebroken gezin (Kosinski *et al.* 2013).

Volgens Eli Pariser worden onze zoekresultaten en levens door deze vergaande verpersoonlijking minder divers. Dit noemt hij de *Filter Bubble*, waarmee hij bedoelt dat internetgebruikers door personalisering in een bubbel terecht komen die ons geeft wat wij willen (Pariser 2011). Als ik altijd klik op roddelsites zal het systeem mij proactief roddelsites voorschotelen en andere sites waar ik niet op klik wegfilteren. De filter bubble bepaalt ons ‘informatiedieet’. “Het consumeren van data die ons wereldbeeld bevestigt is gemakkelijk en plezierig”, schrijft Pariser, “terwijl informatie die ons uitdaagt om

anders over dingen na te denken of info die onze assumpties uitdaagt frustrerend en moeilijk is” (ibid). De auteur voorspelt dat filter bubbels op den duur innovatie en diversiteit remmen.

Een beroemd voorbeeld is opgetekend in de *NY Times* en beschrijft een woedende man die met kortingbonnen een warenhuis binnenstormt. “Mijn dochter kreeg deze toegemild”, zei hij. “Ze zit nog op de middelbare school en jullie sturen haar kortingsbonnen voor babykleren en wiegjes? Proberen jullie haar aan te moedigen zwanger te worden?” (Duhigg 2012). Toen de manager van de winkel dit hoorde belde hij de man op om zijn verontschuldigen aan te bieden, maar inmiddels was de toon anders. “Ze is in augustus uitgerekend. Ik ben u excuses verschuldigd.” Op basis van eerdere aankopen kon niet alleen berekend worden dat deze klant zwanger was maar ook in welk trimester van de zwangerschap ze zich bevond.

Deze *predictive analytics* is een ultieme versie van personalisering, waarbij organisaties op basis van data het vertrouwen hebben dat ze de klant beter kennen dan de klant zichzelf kent. Een serieuze vraag die hierbij opkomt is: wie mag in de toekomst jouw verhaal vertellen? Uitgebreide Big Data-sets over jou (met je browsegeschiedenis, informatie over het boek dat je aan het lezen bent, updates over je relatiestatus) zullen ertoe leiden dat men het gevoel heeft je volledig te kennen. De personalisering die aan de hand van deze data op je wordt toegepast is gedeeltelijk afstemming op persoonlijke wensen, maar heeft ook de mogelijkheid om ons te ‘nudgen’, te overtuigen en te sturen richting dat wat een bedrijf of overheid met ons wil. Het weerhoudt ons hierdoor misschien van onze eigen voorkeuren.

Het stijgend vertrouwen in deze data roept deterministische en dystopische toekomstbeelden op. In de sciencefictionfilm *Gattaca* (1997) bijvoorbeeld, draait het om Big Data-analyses op basis van DNA-profielen. Omdat de hoofdpersoon als genetisch inferieur wordt bestempeld mag hij geen astronaut worden en niet het leven leiden dat hij zou willen. Uiteindelijk slaagt hij erin dit toch te doen (het is een Hollywood film), maar het is een mooie uitvergroting van een dreigend Big Data-probleem. Hoewel selectie op basis van genetische data nu fictie is, is het van belang om te waarborgen dat onze identiteit en onze verlangens niet worden beschadigd door intelligente systemen. Dit is wat Jeroen van den Hoven in zijn visiestuk (p. 43) morele autonomie noemt. Dit betekent dat zelfs als ik met genetische en duizenden andere data tot bepaalde conclusies over jou kom, het aan jou en niet aan mij is om beslissingen te nemen over jouw leven.

Deze kwestie wordt in de toekomst prangender, omdat Big Data niet alleen analyses van ons verleden maakt, maar ook vooruitblijkt en ons adviseert over onze toekomst. Apparaten om ons heen krijgen steeds vaker voorspellende algoritmes mee; onze digitale agenda’s voorspellen wanneer we van huis moeten gaan om op tijd bij de volgende afspraak te zijn en nemen verkeersinformatie, weersinformatie en persoonlijke data

mee in dat advies. Voorspellende televisies zullen op basis van kijkgedrag bepalen welke programma's ze moeten opnemen en slimme koelkasten geven aan wanneer de melk op zal zijn. Sommige systemen gaan verder dan adviseren en koppelen een actie aan een analyse.

Zo heeft Amazon patenten aangevraagd om in de toekomst *anticipatory shipping* in te zetten. Bij deze methode kunnen op basis van aankoopgedrag pakketjes worden afgeleverd voordat de klant op de 'koopknop' heeft geklikt. Als je iedere woensdag een fles olijfolie bestelt, waarom zou dit dan niet een geautomatiseerde aankoop kunnen worden? Natuurlijk zullen er hierbij wel eens fouten worden gemaakt, maar het bedrijf hoopt door middel van kortingen of giften om welwillendheid op te bouwen en vertrouwen te winnen voor dit soort pakketbezorging (Bensinger 2014).

In het rapport *Berekende risico's* waarschuwt het Rathenau dat verzekeraars met het inzetten van data invloed kunnen uitoefenen op de dagelijkse keuzes van klanten. Zo zijn er al verzekeraars die klanten belonen voor een gezonde levensstijl en hen voorzien van fitness-trackers om hun beweging te monitoren. Dit kan direct worden gekoppeld aan financiële prikkels: "een avondje bioscoop levert je minder punten op dan een avondje dansen of pingpongen" (Timmer *et al.* 2015).

Een nieuw profiel

Naast de tech-reuzen uit Silicon Valley zijn er steeds meer datahandelaren die stilletjes in de datamarkt zijn gaan opereren. Er is een miljoenenindustrie ontstaan van *dataprokers* die buiten het zicht van consumenten data over hen verzamelen en deze in vrijwel alle grote sectoren doorverkopen (Senate 2013). Experian, Epsilon en Acxiom zijn de grootste brokers en de laatste kijkt volgens de New York Times actiever en dieper mee in uw leven dan de FBI of de Belastingdienst (Singer 2012). Na de aanslagen op de World Trade Towers in New York bleek dat Acxiom meer wist over 11 van de 19 vliegtuigkapers dan de Amerikaanse overheid. Het bedrijf werkt samen met de National Security Agency (Pariser 2011) en verhandelt data van meer dan een miljard mensen, waarschijnlijk ook van u. Het volgt met een eigen cookie-systeem het onlinegedrag van personen en verzamelt meer dan 5000 data-attributen per persoon (zie figuur 3.2 voor een illustratieve lijst). Hiermee kan toekomstig consumentengedrag worden voorspeld en dit creëert veel meerwaarde voor afnemers.

Op basis van de data-attributen zijn categorieën van consumenten opgesteld, waar iedereen in wordt gegoten. Er zijn zeventig hoofdcategorieën en die variëren van zeer rijke tot werkloze personen. *Shooting stars* zijn bijvoorbeeld kinderloze stellen van gemiddeld 36 met topinkomens, een universitaire opleiding, die veel uitgeven aan reizen, sporten zoals golf en sociale evenementen, restaurants en clubs. *Resilient renters* daarentegen zijn etnisch diverse alleenstaanden die geen huis bezitten, na de middelbare

school niet verder hebben gestudeerd, een zeer laag of geen inkomen hebben, weinig reizen en goedkope kleding kopen. Hoewel deze profielen online staan is het niet algemeen bekend dat bedrijven op deze manier naar ons kijken en ons indelen als een klant die juist wel of juist niet benaderd moet worden. De Zuid-Koreaanse filosoof Han stelt dat deze Big Data nu en vooral in de toekomst leidt tot nieuwe verzuiling; we krijgen zo een digitale klassenmaatschappij. Han stelt dat we naast het bekende gevangenismodel van Bentham (de *panopticon*) nu ook te maken krijgen met een ban-opticon (Han 2015). De panopticon is een ontwerp dat ervoor zorgt dat iedereen in de gevangenis in de gaten wordt gehouden, terwijl de ban-opticon op basis van voldoende data zorgt dat personen die onwenselijk of schadelijk zijn uit een systeem worden gesloten. Zo zal de *resilient renter* misschien minder snel worden benaderd om klant te worden van een bank of verzekeraar dan de *shooting star* en krijgt die laatste misschien zelfs een welkomstpremie.

Bigot Data

Het idee van een banopticon laat zien dat Big Data sociale gevolgen kan hebben die niet zozeer privacygerelateerd zijn, maar die te maken hebben met andere schendingen zoals discriminatie tegen individuen en groepen. Dit hoeft niet bewust te worden gedaan. In Boston heeft de gemeente een app ontwikkeld die op basis van GPS en de accelerometer in een smartphone gaten in de weg detecteert. Doordat honderden gebruikers de app downloaden en meenemen tijdens hun autoritten, krijgt de gemeente accuraat inzicht in de staat van de wegen en de noodzaak bepaalde gedeelten te repareren. Wat het ontwikkelteam in eerste instantie niet had bedacht is dat niet iedereen een auto, smartphone, of bereidheid heeft om deze app te installeren. Hoewel de technologie neutraal is en geen kenmerken van gebruikers meeneemt, kan ongelijke toegang tot technologie ertoe leiden dat er meer signalering en dus ook meer wegverbetering is in welgestelde wijken dan in armere wijken.

Sommige mensen vrezen dan ook dat Big Data zal bijdragen aan de automatisering van discriminatie. Een voorbeeld. Afgelopen jaar ontdekten gebruikers van Google's photo-app (Barr 2015) dat het systeem automatisch zwarte personen als gorilla's categoriseerde. Onderzoekers toonden ook aan dat vacatures voor hoogbetaalde posities via Google minder vaak aan vrouwen worden getoond dan aan mannen. Google bood haar excuses aan, hier zat geen menselijke intentie achter. Facebook kreeg rond diezelfde tijd klachten dat gebruikers met verschillende achtergronden verschillende advertenties te zien kreeg en paste dit beleid – wat bewust was opgezet zodat adverteerders specifieke etnische groepen konden bereiken met 'relevante' reclame voor huizen, banen en kredietaanbiedingen – aan. Facebook heeft inmiddels de advertentietoepassing 'Etnische Affiniteit' uitgeschakeld.

Figuur 3.2 Illustratieve lijst met data die over u verzameld wordt (bron: FTC Rapport)

Identifying Data

Name
 Previously Used Names
 Address
 Address History
 Longitude and Latitude
 Phone Numbers
 Email Address

Sensitive Identifying Data

Social Security Number
 Driver's License Number
 Birth Date
 Birth Dates of Each Child
 in Household
 Birth Date of Family
 Members in Household

Demographic Data

Age
 Height
 Weight
 Gender
 Race & Ethnicity
 Country of Origin
 Religion (by Surname at
 the Household Level)
 Language
 Marital Status
 Presence of Elderly Parent
 Presence of Children
 in Household
 Education Level
 Occupation
 Family Ties
 Demographic Characteristics
 of Family Members
 in Household
 Number of Surnames
 in Household
 Veteran in Household
 Grandparent in House
 Spanish Speaker
 Foreign Language Household
 (e.g., Russian, Hindi,
 Tagalog, Cantonese)
 Households with a
 Householder who is
 Hispanic Origin or Latino
 Employed – White
 Collar Occupation
 Employed – Blue Collar
 Occupation
 Work at Home Flag
 Length of Residence
 Household Size
 Congressional District

Single Parent with Children
 Ethnic and Religious
 Affiliations

Court and Public Record Data

Bankruptcies
 Criminal Offenses and
 Convictions
 Judgments
 Liens
 Marriage Licenses
 State Licenses and
 Registrations (e.g., Hunting,
 Fishing, Professional)
 Voting Registration and
 Party Identification

**Social Media and
 Technology Data**

Electronics Purchases
 Friend Connections
 Internet Connection Type
 Internet Provider
 Level of Usage
 Heavy Facebook User
 Heavy Twitter User
 Twitter User with 250+ Friends
 Is a Member of over 5
 Social Networks
 Online Influence
 Operating System
 Software Purchases
 Type of Media Posted
 Uploaded Pictures
 Use of Long Distance
 Calling Services
 Presence of Computer Owner
 Use of Mobile Devices
 Social Media and Internet
 Accounts including:
 Digg, Facebook, Flickr,
 Flixster, Friendster, hi5,
 Hotmail, LinkedIn, Live
 Journal, MySpace, Twitter,
 Amazon, Bebo, CafeMom,
 DailyMotion, Match,
 myYearbook, NBA.com,
 Pandora, Photobucket,
 WordPress, and Yahoo

Home and Neighborhood Data

Census Tract Data
 Address Coded as Public/
 Government Housing
 Dwelling Type
 Heating and Cooling
 Home Equity
 Home Loan Amount

and Interest Rate
 Home Size
 Lender Type
 Length of Residence
 Listing Price
 Market Value
 Move Date
 Neighborhood Criminal,
 Demographic, and
 Business Data
 Number of Baths
 Number of Rooms
 Number of Units
 Presence of Fireplace
 Presence of Garage
 Presence of Home Pool
 Rent Price
 Type of Owner
 Type of Roof
 Year Built

General Interest Data

Apparel Preferences
 Attendance at Sporting Events
 Charitable Giving
 Gambling – Casinos
 Gambling – State Lotteries
 Thrifty Elders
 Life Events (e.g.,
 Retirement, Newlywed,
 Expectant Parent)
 Magazine and Catalog
 Subscriptions
 Media Channels Used
 Participation in Outdoor
 Activities (e.g.,
 Golf, Motorcycling,
 Skiing, Camping)
 Participation in Sweepstakes
 or Contests
 Pets
 Dog Owner
 Political Leanings
 Assimilation Code
 Preferred Celebrities
 Preferred Movie Genres
 Preferred Music Genres
 Reading and Listening
 Preferences
 Donor (e.g., Religious,
 Political, Health Causes)
 Financial Newsletter
 Subscriber
 Upscale Retail Card Holder
 Affluent Baby Boomer
 Working-Class Moms

Working Woman
 African-American Professional
 Membership Clubs – Self-Help
 Membership Clubs – Wines
 Exercise – Sporty Living
 Winter Activity Enthusiast
 Participant – Motorcycling
 Outdoor/Hunting & Shooting
 Biker/Hell’s Angels
 Santa Fe/Native
 American Lifestyle
 New Age/Organic Lifestyle
 Is a Member of over 5
 Shopping Sites
 Media Channel Usage
 – Daytime TV
 Bible Lifestyle
 Leans Left
 Political Conservative
 Political Liberal
 Activism & Social Issues
Financial Data
 Ability to Afford Products
 Credit Card User
 Presence of Gold or
 Platinum Card
 Credit Worthiness
 Recent Mortgage Borrower
 Pennywise Mortgagee
 Financially Challenged
 Owns Stocks or Bonds
 Investment Interests
 Discretionary Income Level
 Credit Active
 Credit Relationship with
 Financial or Loan Company
 Credit Relationship with
 Low-End Standalone
 Department Store
 Number of Investment
 Properties Owned
 Estimated Income
 Life Insurance
 Loans
 Net Worth Indicator
 Underbanked Indicator
 Tax Return Transcripts
 Type of Credit Cards
Vehicle Data
 Brand Preferences
 Insurance Renewal
 Make & Model
 Vehicles Owned
 Vehicle Identification Numbers
 Vehicle Value Index

Propensity to Purchase a
 New or Used Vehicle
 Propensity to Purchase a
 Particular Vehicle Type
 (e.g., SUV, Coupe, Sedan)
 Motor Cycle Owner (e.g.,
 Harley, Off-Road Trail Bike)
 Motor Cycle Purchased
 0-6 Months Ago
 Boat Owner
 Purchase Date
 Purchase Information
 Intend to Purchase – Vehicle
Travel Data
 Read Books or Magazines
 About Travel
 Travel Purchase –
 Highest Price Paid
 Date of Last Travel Purchase
 Air Services – Frequent Flyer
 Vacation Property
 Vacation Type (e.g., Casino,
 Time Share, Cruises, RV)
 Cruises Booked
 Preferred Vacation Destination
 Preferred Airline
Purchase Behavior Data
 Amount Spent on Goods
 Buying Activity
 Method of Payment
 Number of Orders
 Buying Channel Preference
 (e.g., Internet, Mail, Phone)
 Types of Purchases
 Military Memorabilia/
 Weaponry
 Shooting Games
 Guns and Ammunition
 Christian Religious Products
 Jewish Holidays/ Judaica Gifts
 Kwanzaa/African-
 Americana Gifts
 Type of Entertainment
 Purchased
 Type of Food Purchased
 Average Days Between Orders
 Last Online Order Date
 Last Offline Order Date
 Online Orders \$500-
 \$999.99 Range
 Offline Orders E1000+ Range
 Number of Orders – Low-
 Scale Catalogs
 Number of Orders – High-
 Scale Catalogs

Retail Purchases – Most
 Frequent Category
 Mail Order Responder
 – Insurance
 Mailability Score
 Dollars – Apparel –
 Women’s Plus Sizes
 Dollars – Apparel –
 Men’s Big & Tall
 Books – Mind & Body/
 Self-Help
 Internet Shopper
 Novelty Elvis
Health Data
 Ailment and Prescription
 Online Search Propensity
 Propensity to Order
 Prescriptions by Mail
 Smoker in Household
 Tobacco Usage
 Over the Counter
 Drug Purchases
 Geriatric Supplies
 Use of Corrective Lenses
 or Contacts
 Allergy Sufferer
 Have Individual Health
 Insurance Plan
 Buy Disability Insurance
 Buy Supplemental to
 Medicare/Medicaid
 Individual Insurance
 Brand Name Medicine
 Preference
 Magazines – Health
 Weight Loss & Supplements
 Purchase History or Reported
 Interest in Health Topics
 including: Allergies,
 Arthritis, Medicine
 Preferences, Cholesterol,
 Diabetes, Dieting, Body
 Shaping, Alternative
 Medicine, Beauty/Physical
 Enhancement, Disabilities,
 Homeopathic Remedies,
 Organic Focus, Orthopedics,
 and Senior Needs

Onderzoekssite Propublica ontdekte een ernstiger probleem bij software die gebruikt wordt om het risico op recidivisme in de justitiële sector te voorspellen. Op basis van Big Data-analyses maakte het systeem de onjuiste voorspelling dat bepaalde zwarte beklagden een hoger risicoprofiel hadden om te recidiveren (Angwin *et al.* 2016). Blanke beklagden werden onjuist een laag risicoprofiel toegekend. Dit zijn data-problemen. Algoritmes leren op basis van de foto's, teksten, en data waarmee ze worden gevoed. Als ingenieurs een set data met een bias erin voeden aan een systeem, dan bouwt het systeem een model van de wereld dat hierop gebaseerd is.

De oplossing is om meer complete en neutrale datasets te gebruiken. Maar wat nu als er vooroordelen zitten in de grootste datasets? Als Big Data en KI leren van onze huidige wereld bestaat het risico dat ze ook het onrecht, de ongelijkheid, en de problemen van onze wereld overnemen en internaliseren. Een voorbeeld dat stof deed opwaaien en tot veel hilariteit leidde was Tay AI. Tay zag het licht op 23 maart 2016 en was door Microsoft ontworpen om geheel autonoom en gebaseerd op het taalpatroon van een 19-jarig Amerikaans meisje op twitter gesprekken te voeren en te leren van online menselijke interactie. Na het initiële "hello world" sloeg ze al snel een minder vriendelijke toon aan:



Tay AI en haar 'Bigot Data' werden binnen zestien uur uitgezet. Per ongeluk werd ze op 30 maart 2016 nog eens geactiveerd, maar na gelijksoortige seksistische en racistische tirades ging het licht definitief uit.

Bij de inzet van voorspellende data om te bepalen hoe politie wordt ingezet of hoe lang iemand naar de gevangenis moet, is het van groot belang om te weten welke data

gebruikt worden en of de historische data niet leiden tot een ongewenste toekomst. Big Data heeft geschiedenis en context en dit kan ertoe leiden dat onjuistheden uit het verleden doorleven in de digitale platforms van de toekomst. De intransparantie van systemen maakt het moeilijk om deze algoritmische fouten te ontdekken. Hoe kom je er als buurtbewoner achter dat software bepaalt dat er minder gepatrouilleerd wordt in jouw straat? Hoe solliciteer je op een baan die je nooit onder ogen hebt gekregen?

Privacy

“Privacy is diefstal. Delen is liefde. Geheimen zijn leugens”. De motto’s van het Orwelliaanse bedrijf The Circle uit de gelijknamige roman van Dave Eggers zijn duidelijk; we moeten het privacybegrip afschaffen omdat het radicale transparantie in de weg zit (2014). De meningen in de privacydiscussie lopen sterk uiteen. Net als The Circle voorzien sommige technologiebedrijven een toekomstscenario waarin het privacybegrip verschuift en minder relevant wordt. Zo stelde bestuursvoorzitter Eric Schmidt van Alphabet (het moederbedrijf van Google) in een televisie-interview op CNBC: “Als je iets doet waarvan je wilt dat niemand het weet, zou je het misschien helemaal niet moeten doen” (Esguerra 2009). Daar tegenover staat iemand als politiek activist Peter Handke die schrijft: “Wat anderen niet van mij weten, daar leef ik van.” (In Han 2015).

Privacy is een onderwerp dat bij uitstek centraal staat bij de opkomst van nieuwe informatietechnologieën. De eerste argumenten om privacy in de wet op te nemen – rond 1890 door Warren en Brandeis in de Verenigde Staten – hadden te maken met de opkomst van nieuwe communicatietechnologie zoals de reproductie en verspreiding van foto’s in nationale kranten en tijdschriften. De vraag is echter waar we het bij privacy precies over hebben. Solove stelt dat het een concept is dat in wanorde verkeert, waarvan niemand precies kan articuleren wat het inhoudt. Een krant die de naam van een slachtoffer publiceert en een bedrijf dat een lijst maakt van vijf miljoen incontinente vrouwen begaan allebei een privacy-schending, maar het is duidelijk niet dezelfde soort schending (2006). Het begrip privacy is een breed begrip dat gaat over de bescherming van persoonsgegevens, de bescherming van het eigen lichaam en van de eigen woning, de bescherming van familie- en gezinsleven en het recht om vertrouwelijk te communiceren. Met het oog op Big Data en de toekomst zien we een aantal ontwikkelingen waarin de intrinsieke eigenschappen van Big Data schuren met het privacybegrip (Working Group on Data Protection in Telecommunications 2014).

Ten eerste rekt Big Data het begrip persoonsgegevens op omdat men aan de hand van miljoenen data-elementen nieuwe profielen kan maken en tot nieuwe manieren kan komen om personen te identificeren. Door middel van analyses kunnen – op zichzelf onschuldige – gegevens na het verwerken, combineren en analyseren gevoelige resultaten opleveren, bijvoorbeeld over iemands gedrag, voorkeuren of verhoogde risico’s.

Ten tweede heeft Big Data ‘honger’; de meerwaarde van Big Data komt namelijk voort uit het verzamelen en verwerken van zoveel mogelijk informatie. Deze data-honger staat op gespannen voet met het in Europa populaire beleidsconcept van data-minimalisatie, dat stelt dat alleen minimale en vooraf bepaalde data mogen worden verzameld en geanalyseerd. Doelbinding als principe wordt met het verzamelen van Big Data ook vaak overboord gegooid. Big Data heeft betrekking op het hergebruik van informatie voor nieuwe doelen, die soms heel anders zijn dan de doelen waarvoor de informatie verkregen is.

Hier komt nog bij dat er bij veel analyses een gebrek is aan transparantie; niet alleen gebeurt het analyseren van data niet openlijk, het gebeurt in grote mate geautomatiseerd zodat het voor mensen moeilijk navolgbaar is.

Black box society. Over Big Data en transparantie

In 1985 sprak filosoof en computer-ethicus James Moor in zijn artikel ‘What is Computer Ethics?’ zijn bezorgdheid uit over de onzichtbaarheidsfactor van computertechnologie. Die zou volgens hem de deur openzetten voor heimelijke controle, computercriminaliteit en privacyinbreuken. De technologie kan immers zo geprogrammeerd worden dat de gebruiker niet eens doorheeft dat men met gegevens aan de haal gaat of stiekem meegluurt. Dertig jaar later werd de bezorgdheid herhaald: op een mediaconferentie in München in 2016 stelde Angela Merkel dat “algoritmes transparanter moeten worden gemaakt” (Connolly 2016). Ze was van mening dat geïnteresseerde burgers zich moeten kunnen informeren over dat wat hun gedrag en dat van anderen online stuurt. “Algoritmes kunnen, als ze niet transparant zijn, onze percepties verstoren en onze kennis doen krimpen”, vervolgde ze. “Grote internetplatforms met hun algoritmes en filters zijn het oog van de naald geworden waar talloze media doorheen moeten kruipen [om hun gebruikers te bereiken].”

Dit is de Transparantie Paradox. Big Data leidt enerzijds tot een wereld die meer transparant is en wereldwijde kennis binnen handbereik brengt, maar zorgt er anderzijds voor dat het verzamelen van data onzichtbaar gebeurt en dat de instrumenten die ermee gemoeid zijn vaak zowel fysiek, technisch, als juridisch zijn afgeschermd. Corien Prins en Lokke Moerel van de Universiteit Tilburg maken een soortgelijke observatie en schrijven dat het individu meer transparant wordt voor iedereen, behalve voor zichzelf (2016). Ze stellen dat individuen nu vaak niet weten dat er data over hen worden verzameld, hoe deze worden verwerkt, door wie de data worden opgeslagen en welke gevolgen het analyseren ervan voor hen kan hebben.

Er zijn wel allerlei beleidsmaatregelen die stellen dat het individu geïnformeerd moet worden of dat er toegang gevraagd moet worden, maar in de praktijk zijn deze maatregelen niet altijd goed uitgewerkt. Zo worden voorwaarden van data-applicaties

zeer gebruikersonvriendelijk opgesteld en (daardoor) niet gelezen. Onderzoekers van de Carnegie Mellon Universiteit berekenden dat het een gemiddelde internetgebruiker 76 werkdagen per jaar zou kosten om de privacymaatregelen te lezen van de gangbare websites en apps die men gebruikt (Sadeh *et al.* 2014). Dit komt erop neer dat je vijftien weken per jaar de hele dag privacymaatregelen moet lezen om er zeker van te zijn dat er geen kleine lettertjes zijn die jou kunnen schaden.

In de transparantie-discussie gaat het behalve over intransparante regels vooral over het ontwerp van systemen. In het vorige hoofdstuk omschreven we een toekomstbeeld waarin Ambient Intelligence bepalend is voor de inrichting van onze publieke en private ruimte. We werken deze momenteel al uit met ideeën zoals het Internet of Things, en de ontwerpprincipes zijn vrijwel gelijk aan de voorwaarden uit de ISTAG scenario's (Ducastel 2001).

- Eis 1: Zeer onopvallende hardware
- Eis 2: Een naadloos schakelende communicatie-infrastructuur
- Eis 3: Dynamisch en massaal gedistribueerde apparaatnetwerken
- Eis 4: Natuurlijk gevoel bij de interfaces
- Eis 5: Betrouwbaar en veilig

De sensoren die we ontwerpen om data te verzamelen worden zo gebouwd dat ze zeer onopvallend zijn, en zich naadloos én natuurlijk manifesteren in onze leefwereld. Deze principes sluiten aan bij de gewenste gebruikerservaring van klanten, maar camoufleren ook actief hoe systemen werken, wat er verzameld wordt en wie er meekijkt. Martijn en Tokmetzis schetsen het beeld van een grote groep mannen die je achtervolgen tijdens het winkelen en meekijken hoe lang je naar een product kijkt, wat je koopt, met wie je contact hebt en waar je woont. Intimiderend? Eng? Dit is de dagelijkse realiteit van wat er online gebeurt nu cookies van talloze organisaties je digitale reilen en zeilen op de voet volgen (Martijn en Tokmetzis 2016).

De *Black Box society* is een toekomstbeeld dat vandaag de dag al ervaren wordt. Bij veel van onze apparatuur zien we wat het systeem in gaat en wat eruit komt, maar snappen we niet wat er binnenin gebeurt. Maar waar we bij televisies geen behoefte hebben om precies te begrijpen hoe energie wordt omgezet in beelden, of bij auto's niet hoeven te weten hoe een verbrandingsmotor exact werkt, is er bij datasystemen meer zorg om de exacte werking van de algoritmes. De zwarte doos is vaak zwarter dan bij andere technologie, omdat ook de exacte input en output onbekend is voor de gebruikers. Daarnaast is de output van een heel andere aard. De systemen zorgen niet voor vermaak of transport, maar simuleren intelligentie en leiden tot analyses, adviezen of beslissingen die gevolgen kunnen hebben voor onze levens.

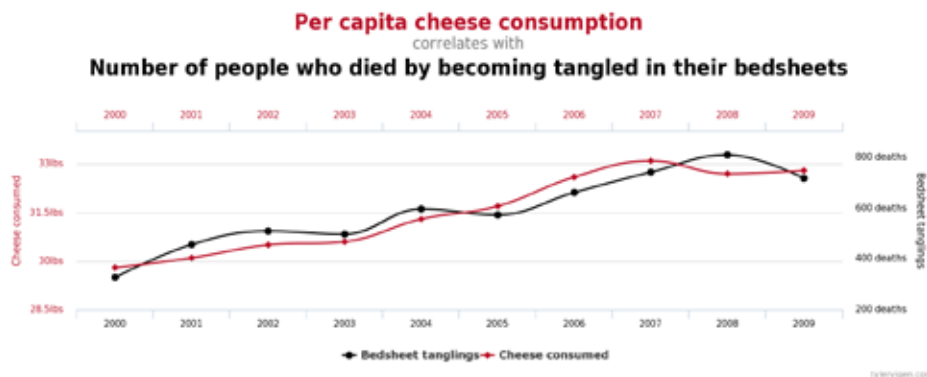
Een Big Data-systeem kan op basis van input bepalen of iemand geschikt is voor een baan, fraudeert met zijn belastingaangifte, of een opleiding binnen de voorgeschreven tijd kan behalen. Het systeem werkt als een lens die helpt bepalen hoe een organisatie zoals een overheid naar jou kijkt en hoe ze over jou oordeelt. Als jouw verzekeringspolis op basis van een data-analyse duurder blijkt te zijn dan die van je buurman, dan wil je zeker weten dat de gebruikte informatie juist, volledig en van goede kwaliteit is. Daarnaast wil je waarschijnlijk weten hoe de afweging tot stand komt: is je leeftijd een bepalende factor voor je premie, of is het je opleiding, je dieet, of je wekelijkse hardlooptijd?

Crawford en Schultz schrijven in de context van privacy en Big Data dat de uitkomsten van slimme datasystemen vaak niet alleen onvoorspelbaar of onbegrijpelijk zijn voor gebruikers, maar in de toekomst ook steeds meer voor hun programmeurs (2014). Privacy wordt hierdoor problematisch om te waarborgen, omdat men niet kan voorspellen welke persoonskenmerken een zelflerend systeem zal achterhalen of welke maatregelen men preventief zou moeten nemen om inbreuk op privacy te voorkomen. Dit geldt echter ook in breder verband. De uitkomsten van slimme datasystemen worden meer complex en minder voorspelbaar. Algoritmische transparantie en inzage in de werking van dergelijke systemen wordt daarmee een grote uitdaging.

Een belangrijke vraag hierbij is ook: transparantie voor wie? Als een systeem een complexe analyse uitvoert en hier duizenden data-elementen in meeneemt, heeft de eindgebruiker dan baat bij een overzicht van alle stappen en data die het systeem meeweegt? Of is dit even onbruikbare informatie als de te lange privacyvoorwaarden van onze apps?

In data we trust

Een risico van een black-box-samenleving is dat we misplaatst vertrouwen kunnen hebben in technologie. Journalist en onderzoeker Timandra Harkness waarschuwt ervoor dat dit gevaarlijk kan uitpakken. “We vertrouwen onszelf niet meer. Dat begint al bij het ontbijt: eet ik wel gezond genoeg, maak ik de aarde niet kapot als ik dit worstje eet? We weten

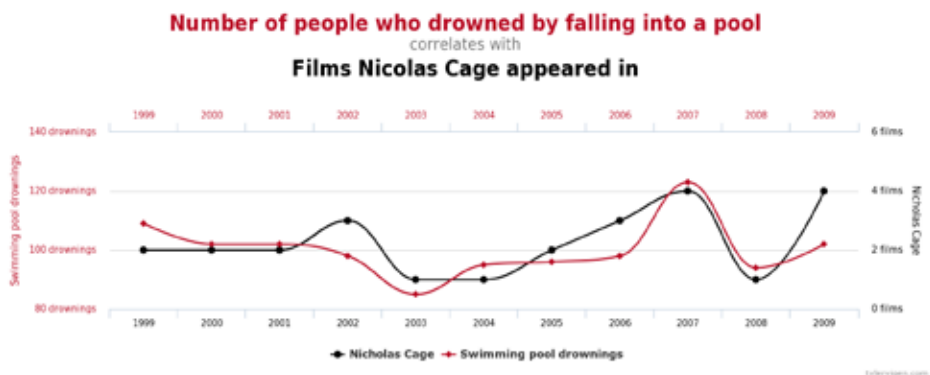


het niet”, stelt ze in een interview (Visscher 2016). Omdat we geen vertrouwen hebben in ons eigen oordeel en ook experts wantrouwen vanwege mogelijke belangenverstrengelingen kijken we uit naar Big Data-systemen. Deze hebben een schijn van wetenschappelijkheid en accuraatheid omdat alles kan worden uitgedrukt in exacte getallen. “Big Data wekt de suggestie dat het een compleet beeld van ons leven kan schetsen, maar er zijn nu eenmaal zaken die niet zijn uit te drukken in cijfers”, stelt Harkness.

Onlangs is onderzoek gedaan onder dertig personen bij het Georgia Institute of Technology in Atlanta om ons vertrouwen in technologie te toetsen (Ruthkin 2016). Een student zit op een kamer met een robot en voert een test uit als het rookalarm afgaat. De student moet kiezen: of hij ontsnapt via de manier waarop hij binnenkwam – een route die duidelijk staat aangegeven – of hij volgt de richting die de robot aanwijst, een obscuur gangetje dat de student niet kent. Van de dertig studenten volgden 26 studenten de robot, bleven twee studenten in de kamer, en werden twee studenten uit de studie verwijderd vanwege ongerelateerde redenen.

Het vertrouwen in technologie is klaarblijkelijk groot. Sherry Turkle schrijft in haar boek *Alone, together?* dat we hierdoor steeds minder verlangen van andere mensen. We krijgen volgens haar de illusie dat we sociaal verbonden zijn zonder dat we met andere mensen omgaan (Turkle 2011). Misschien raken we echter meer sociaal verbonden met technologie? Dit zou ons oproepen om onze sociale verhouding tot technologieën te onderzoeken en ontwikkelen. We zullen ons traditionele begrip van vriendschap, kameradschap of relaties moeten uitbreiden om ook met kunstmatige intelligentievormen relaties aan te kunnen gaan.

Tegelijk moeten we ons ervan bewust zijn dat data niet gelijk staat aan waarheid. “Martel de data, en het zal alles wat je maar wilt bekennen”, zei Ronald Coase, Nobelprijswinnaar in de economie, al. Big Data blijft multi-interpretabel. Dit blijkt ook uit een data-verschijnsel dat apofenie genoemd wordt, waarbij we in data willekeurige patronen kunnen opmerken en daar betekenis uit kunnen afleiden. Tyler Vigen creëerde een website met dergelijke *spurious correlations*.



Big Data is niet altijd betere data; en kwantiteit betekent niet automatisch kwaliteit. Problemen die we met 'kleine data' hebben op het gebied van datakwaliteit (is de data accuraat, precies, consistent, is de bron bekend?) gelden bij het binnenhalen van Big Data op grote schaal evenzeer. Neem bijvoorbeeld data van sociale media – tweets, likes, apps – die steeds meer worden meegewogen om de stemming in het land te peilen. Het is erg lastig op basis van miljoenen tweets uitspraken te doen over wat mensen vinden. Bovendien spelen er allerlei dataproblemen; de Twitter-populatie is niet representatief voor de wereldbevolking, veel tweets zijn niet serieus, sommige gebruikers zijn bots, en bronnen zijn slecht te achterhalen of onbekend (boyd en Crawford 2011).

Over Big Data en machtsverhoudingen

Een studie in *Science Magazine* omschreef afgelopen jaar het 'zoekmachine-manipulatie-effect' (Shultz *et al.* 2015). Omdat zoekmachines zo goed zijn geworden in het rangschikken van zoekopdrachten, heeft de plaats op de ranglijst een groot effect op het vertrouwen van gebruikers. Als je zoekt naar het beste boek over Big Data zul je eerder bij de eerste paar bevindingen uitkomen dan bij een boek dat pas wordt genoemd op pagina 4. Dit is bekend, maar de vraag is of dit principe ook zou kunnen werken in politieke processen zoals de nationale verkiezingen. Auteur van de studie, Robert Epstein, deed onderzoek onder kiezers rond de nationale verkiezingen van India in 2014 en kwam tot de conclusie dat het manipuleren van zoekopdrachten door kandidaten hoger of lager te plaatsen een significant effect had. De groep zwevende kiezers verkoos in deze studie met meer dan 12% voorkeurstemmen de omhoog gemanipuleerde kandidaat.

Een ander onderzoek demonstreerde dat Facebook mensen succesvol kan mobiliseren om te stemmen door herinneringen te plaatsen op hun tijdslijn. Alhoewel er geen aanwijzingen zijn dat tech-bedrijven zich actief mengen in verkiezingen en via hun systemen voorkeur uitspreken voor kandidaten, bestaat de mogelijkheid wel. Op dit moment is het zo dat algoritmes van bedrijven als Google of Facebook als het geheim van de smid zijn. Op basis van concurrentiebeding en intellectuele eigendomsrechten kunnen bedrijven hun algoritmes privé houden. De mogelijkheid tot manipulatie is echter een vorm van machtsmisbruik die op de loer ligt bij gesloten Big Data-systemen.

Er zit een bepaalde inbalans in de toepassing van dit soort technologie. Big Data geeft kennis en dus macht aan diegene die toegang heeft tot de technologie. Morozov schreef hierover dat het niet alleen disruptief is, maar ook de status quo kan beschermen. Dit zijn vaak de grote organisaties, zoals overheden en bedrijven. Het opzetten van Big Data-analyses is immers een kostbare aangelegenheid, waarvoor zowel technologie als mankracht, kapitaal en expertise nodig is. Een kleine partij of de individuele burger kan dit niet opbrengen en wordt zo feitelijk buitenspel gezet. Big Data creëert winnaars en verliezers, maar het is momenteel voor de hand liggend dat individuen het onderspit

delven tegenover grote instituties die het systeem ontwikkelen, de data mijnen en analyseren, en het sorteren van data zelf inrichten. Het is van belang deze onbalans recht te trekken door de positie van de burger of eindgebruiker centraal te stellen in het ontwerp en beheer van datagedreven technologie.

Conclusies: data van waarde

Big Data heeft een hoge economische waarde. Het wordt gezien als een grondstof voor intelligentie en de grootste verdienste van Big Data is dan ook dat het leidt tot intelligentere handelingen, analyses en besluiten. Hierdoor kunnen organisaties beter inzicht krijgen in hun processen dan ze voorheen hadden. Erik Brynjolfsson van MIT stelde vast dat bedrijven die een datagedreven strategie toepassen 5-6 procent meer productief zijn dan bedrijven die dit niet doen, zelfs na aftrek van investeringskosten zoals IT (In Kennedy 2016). McKinsey berekent dat er wereldwijd honderden miljarden omgaan in de Big Data-industrie en stelt dat er vijf manieren zijn waarop organisaties waarde kunnen creëren met Big Data (Chui *et al.* 2010):

1. Transparantie: inzicht krijgen om efficiëntie op te stuwen
2. Analyseren: een meer nauwkeurige analyse en evaluatie van medewerkers en systemen
3. Categoriseren: het segmenteren van populaties om acties op maat te ondernemen
4. Automatisering: ondersteunen en vervangen van menselijke besluitvorming door geautomatiseerde algoritmes
5. Innovatie: het innoveren van nieuwe bedrijfsmodellen, producten en diensten.

Naast deze economische waarde zal datagedreven technologie ook een impact hebben op onze maatschappelijke waarden. Zoals Lessig schrijft, worden we meer en meer gereguleerd door digitale infrastructuur en tast deze techno-regulering onze vrijheden aan (Lessig 2009). Nu we meer tijd besteden in het digitale domein, en meer van onze levens in data wordt gevat is deze digitale infrastructuur zo onzichtbaar geworden dat we hier niet vaak bij stilstaan. De datasystemen en infrastructuur bepalen echter hoe we met privacy omgaan, hoe onze digitale identiteit gevormd wordt, of data door iedereen kunnen worden ingezien of dat ze afgeschermd worden. Het bepaalt wie wat ziet en wat er gemonitord wordt. Kortgezegd hebben datasystemen en de code die deze sturen een steeds grotere invloed op onze vrijheid van meningsuiting, gelijkheid, privacy, en andere basale waarden in onze samenleving. Het is daarom belangrijk als samenleving te verkennen hoe we willen omgaan met deze technologie, en hoe we dit niet willen.

Leven met robo's

Vanmorgen werd ik wakker van de geur van koffie, gemalen door mijn gepersonaliseerde zorgrobot. Het ding heet Ibu en weet mij dag in dag uit te verleiden om de nieuwe dag – met zin of tegenzin – in te gaan. Het respecteert mijn humeur, begrijpt dat wij mensen de oude dag met gebreken bereiken en doet mij met de radicale humor waar ik zo van hou regelmatig hardop lachen. Ibu reciteert gedichten, leest voor uit de NRC-blogosfeer, vertaalt de Chinese berichten van mijn goede vriend uit China, doet de was, masseert mijn voeten, onderhandelt met het slimme huis over mijn energieverbruik en beheert mijn blockchain-accounts zodat ik de juiste *smart contracts* afsluit en financieel overeind blijf. Ibu wordt scherp gehouden door Alim, mijn andere huisgenoot, die door een andere serviceprovider is ontworpen en die even vergaand gepersonaliseerd is, maar dan toch net anders dan Ibu. Door mijn oude dagen met hen beiden door te brengen blijf ik mij bewust van de productieve bias die dit soort systemen genereren en kan ik mij permitteren ze zo nu en dan tegen elkaar uit te spelen. Zodat ik beter zicht krijg op mijn persoonlijke ruimte, de voor mij ontworpen *probability space* of *hypothesis space*, zoals Kolmogorov of Bayes zouden zeggen.

We leven in spannende tijden. Vroeger (zeg 2017) bestond de wereld uit mensen, dieren en planten, aangevuld met instituties (onderwijsinstellingen, kerken, banken, ziekenhuizen, bedrijven, overheden), die vaak een eigen leven leidden met een eigen dynamiek. Nu (2047) hebben we bovendien van doen met allerlei robo's die net als mensen, dieren, planten en instituties hun gedrag aanpassen op basis van feedback: lerende kunstmatige systemen zoals zelfrijdende auto's, zorgrobots, het besturingssysteem van het slimme huis, en onze persoonlijke digitale assistent die permanent in gesprek is met alle andere slimme systemen om de omgeving voor ons te optimaliseren. Op basis van onze gedragsgegevens voorzien al die robo's wat we morgen of volgend jaar doen, waar we dan zijn, wat we willen bereiken en met wie we zullen omgaan. Gelukkig klopt dat meestal niet, omdat we hebben geleerd om in te spelen op al die voorspellingen en daar ons voordeel mee te doen. Het gaat er niet om de toekomst te voorspellen maar om die mogelijk te maken, en dan graag op een veilige manier die ons ruimte geeft nieuwe wegen in te slaan en die de kansen en risico's op een eerlijke manier verdeelt.

Vroeger dachten we dat wij robo's konden

gebruiken alsof het passieve dingen zijn die geen eigen inbreng hebben. Later begonnen we te beseffen dat het misschien wel omgekeerd is: om goed te kunnen inspelen op hun omgeving en ons beter van dienst te zijn hebben robo's onze gedragsgegevens nodig (ze luisteren naar wat we doen, niet naar wat we zeggen te doen). Nog later begrepen we dat ze ons machinaal leesbare gedrag gebruiken om het verdienmodel te voeden van degene die hen aanbiedt (de serviceprovider). Dus: wij gebruiken robo's en zij gebruiken ons, en de serviceprovider gebruikt ons allebei.

Toen dat doordrong kwam er een beweging op gang om beter zicht te krijgen op de

verbanden en het soort verbanden dat wij zelf voortdurend leggen. Je hoeft daarvoor geen wiskundige of computerwetenschapper te worden, maar je moet wel snappen dat niet ieder wiskundig verband interessant of relevant is. Je moet als het ware een gezonde weerstand ontwikkelen tegen al die voorspellingen en een steekhoudend weerwoord kunnen geven als je getarget wordt door je robo (of door die van een ander). Nu leer je op school dus niet alleen het schrijven en lezen van tekst, maar ook het schrijven en lezen van computercode en de statistische verbanden die onze slimme robo's genereren. Lezen en schrijven is een kwestie van oefenen. Als je het

Vroeger dachten we dat wij robo's konden gebruiken alsof het passieve dingen zijn die geen eigen inbreng hebben.

manier waarop al die verschillende robo's ons beïnvloeden. Dat bleek nog niet eenvoudig. Om te beginnen claimden bedrijven hun bedrijfsgeheim, die wilden hun *competitive advantage* niet kwijt. Daarnaast beschermen bedrijven hun intellectuele rechten, omdat ze nu eenmaal veel investeren in al die slimme systemen en die investering moet kunnen renderen. De werking van robo's werd dus opzettelijk verhuld. Los daarvan is die werking gebaseerd op wiskundige verbanden tussen data die ons niet veel zeggen, wij denken niet in termen van wiskundige verbanden. Het heeft even geduurd, maar uiteindelijk bleek dat het heel belangrijk is om te kunnen omgaan met het verschil tussen wiskundige

eenmaal in je vingers hebt kun je ermee gaan spelen en krijg je toegang tot allerlei informatie.

Het mooiste is dat we hebben geleerd om met onze robo's samen te leven. We geven ze het voordeel van de twijfel als ze zich voordoen als invoelende soortgenoten. We weten dat het geen lotgenoten zijn – dat ze geen pijn voelen, geen bewustzijn hebben, laat staan zelfbewustzijn. Maar we weten ook dat ze zich voeden met onze data en voortdurend bijleren. We hebben ontdekt hoe we ze kunnen bijsturen door de manier waarop we met ze omgaan. En door ze tegen elkaar uit te spelen hebben we *checks & balances* en transparantie ingebouwd. Nu ga ik aan de koffie, met Ibu en Alim.

Hossegor 2036, Atlantische kust

Ik heb het je al vaker verteld en ik kan er nog steeds boos om worden. De tweede kredietcrisis was desastreus voor Europa. De economie stortte volledig in. Er waren diverse ernstige conflicten met permanente oorlogsdreiging. De mogelijkheden om de economie verder aan te zwengelen waren uitgeput en door diverse lidstaten is op grote schaal financiële steun van China geaccepteerd. Om deze steun te krijgen hebben we Chinese datagedreven services moeten inkopen. Door de inzet van deze sterk gepersonaliseerde diensten is ook de escalatie van conflicten binnen Europa afgewend, met dank aan de grootste rekenkracht ter wereld en de opgedane

ervaring met gedragsanalyse van de volledige Chinese bevolking. De combinatie van de meest omvangrijke *knowlegde graph* met triljoenen feiten en de allerbeste voorspellingsmodellen van o.a. weer, logistiek, planologie en met name ons persoonlijk gedrag, heeft China wereldleiderschap in alle denkbare diensten bezorgd. Het valt niet te ontkennen dat China alle diensten domineert, van stedenbouwkundig advies en bankzaken tot amusement. China levert nou eenmaal de beste aansluiting op onze directe en toekomstige behoeften. En ja,

wij zijn daar volledig afhankelijk van geworden. Daar staat gelukkig tegenover dat we het economisch inmiddels beter hebben en ik me weer een surftrip naar Hossegor in Zuid-Frankrijk kan veroorloven.

Na een reis van drie uur met de Tube sta ik al op het strand. Een mist van koele druppeltjes slaat neer op mijn wangen. Op de achtergrond hoor ik het dreunen van de branding. De golven zijn meer dan vijf meter hoog. Eigenlijk veel te hoog voor mannen van bijna tachtig. Als we de aanbeveling van de Baibu Alibaba Tencent (BAT) SurfSafetyService geloven, moet het

“Uw hartslag en bloeddruk zijn te hoog en daarom is de kustwachtdrone ingeroepen.”

kunnen. Na mijn heupoperatie van vorige week loop ik een beetje stijf door het zand. “Vincent gaan we er hier in? De *shore break* is best flink,” roep ik. “De surfboards zijn aangedreven,” schreeuwt Vincent over het gebulder van de branding heen. Vincent is al ruim vijftig jaar mijn surfmaatje. “Hopelijk is deze SurfSafetyService niet zo truttig als alle andere BAT-diensten,” zeg ik lichtelijk bezorgd. “Tja, eigen schuld, de BAT-services zijn net zo truttig als ons bejaardengedrag waarop ze zijn gebaseerd,” antwoordt Vincent. “Hm, met dank

aan de vergrijzing,” zeg ik, en we springen door de drie meter hoge shore break heen en peddelen naar de achterliggende golven. Direct verschijnt er een projectie vanuit de BAT-drone. Een vriendelijke BAT-dame zegt met een zeer heldere stem: “Ik raad u aan om 1200 meter naar rechts in zee te gaan.” Ik antwoord: “Hier is een sterke stroming richting zee, daar kunnen we prima gebruik van maken.” “Willen jullie onverzekerd doorgaan?”, reageert de dame. “Ja!”, schreeuwen Vincent en ik geïrriteerd terug. We laten ons door de stroming naar de achterste golven voeren. Ver achter de branding is het altijd wonderbaarlijk rustig en kunnen we wachten op een goede golf. Terwijl we rustig op en neer deinen, stuurt de BAT-dame Vincent naar een mooie cleane golf. Vincent suist met hoge snelheid op het strand af. “Let op de shore break,” denk ik. Een domme gedachte, want voordat de omslaande golf Vincent op het strand neerkwakt, heeft het board hem naar een gemakkelijker plek om te stranden gebracht. Langzaam peddel ik naar een plek waar ik een serie grote golven denk te zien. Eerder dan ik verwacht, komt een muur van water omhoog. Ik probeer er nog onderdoor te duiken. Te laat. Ik neem snel een ademteug. De watermassa duwt me tollend naar de bodem. Ik voel de druk op mijn oren. Terwijl het board me door het water trekt, probeer ik me omhoog te werken. Eindelijk weet ik mijn hoofd boven het kolkende schuim te krijgen: lucht! Er komt nog zo’n grote golf aan. Ik duik snel weer onder en wordt weer genadeloos gespoeld. Prachtig! Zo moet surfen zijn; de angst-kick of je het gaat redden. Ik worstel me weer boven en hoor de BAT-dame zeggen: “Uw hartslag en bloeddruk zijn

te hoog en daarom is de kustwachtdrone ingeroepen.” Nee hè, denk ik, en tegelijk ik zie ik een ongecontroleerd heen en weer bewegende kustwachtdrone. De BAT-dame meldt: “Excuses, de kustwachtdrone kan u niet uit het water tillen. Het zicht is defect.” Ik glimlach en weet al wie hier achter zit. Met een totaal andere stem zegt de BAT-dame dan: “Graag gedaan opa.” Ik herken direct de stem van Erlo, mijn kleindochter. Erlo heeft het BAT-systeem even gehackt. Zij heeft een bedrijf dat improvisatietheater combineert met het ‘creatief’ veranderen van officiële BAT-diensten. Laatst heeft ze een stel van die suffe zelfrijdende autootjes gehackt en een feestje gegeven op een leegstaand vliegveld. Een soort dance battle met autootjes, waarin Erlo de centrale choreografie had en er prijzen te winnen waren voor de beste solo-performances. Onder de jongere generatie is het een ware rage, ontstaan omdat de jeugd nauwelijks een rol heeft in de ontwikkeling van deze BAT-diensten en zij deze diensten nog truttiger vinden dan ik. De rebellerende jeugd past wereldwijd op grote schaal dit soort services aan. Aangezien deze ‘aangepaste’ diensten flinke werkgelegenheid creëren en de overheden angst hebben voor een grootschalige jeugdopstand, worden deze activiteiten oogluikend toegestaan. Dit soort bedrijfjes zijn regionale piraten met onvoorspelbaarheid als handelsmerk. Met ironie realiseer ik me dat de privacywetgeving uit mijn tijd heeft gefaald en dat jongeren op deze manier hardhandig zelfbeschikking terugpakken door zich niet aan de regels te houden. Opa is trots, want Erlo is zo’n piraat die ondanks haar drukke bestaan toch nog even tijd heeft om haar opa’s surfsessie spannender te maken.

China datamodel

De omwenteling van een planeconomie naar een markteconomie bracht veel teweeg in China. Wat blijkt? Als je mensen de ruimte geeft doen ze soms ook dingen die niet de bedoeling zijn, zeker in een land met zwakke instituties. Giftig melkpoeder verkopen, rotzooien met eten, financiële fraude; China heeft recentelijk heel wat schandalen gekend. In de toekomst moet dit anders en daarom is de regering bezig op basis van Big Data een landelijk sociaal kredietsysteem op te zetten. Dit systeem bepaalt aan de hand van data van verschillende partijen kredietscores. Bijzonder hierbij is dat burgers zelf meebouwen aan het scoresysteem door hun mening te geven over artsen, onderwijzers, bedrijven en ambtenaren. Tot op zekere hoogte bestaat het nu al voor bedrijven; Baidu beheert voor de staat een openbare database waarin vooral negatieve profielen zijn opgenomen. Je kunt op deze website bijvoorbeeld zien of bepaalde partijen veroordeeld zijn wegens handel zonder vergunning, fraude, etc.

Wat het meeste stof doet opwaaien is het idee dat je burgers een score kunt toekennen. Ben jij een zesjestype of verdien je een 10? Vanaf 2020 moeten volwassen burgers naast hun identiteitskaart zo'n kredietcode hebben, alhoewel ik moet benadrukken dat de plannen grotendeels nog onduidelijk zijn. Er is een beleidsdocument waarin staat dat je geen vliegtickets of luxe hotelkamers meer

zou kunnen boeken als je wordt aangemerkt als 'onbetrouwbaar'. Mogelijk wel een staanplaats achterin de trein, maar niet in de eerste-klas-coupé.

De databases worden samen met marktpartners gemaakt, zoals Alibaba die met haar Sesame Credit al een unieke database met klantscores heeft. Met deze scores kun je sparen voor prijzen net als met airmiles, maar prijzen zouden ook overheidsdiensten kunnen zijn, zoals een versnelde visumaanvraag voor buurlanden. Miljoenen klanten gebruiken de scores zelf op datingsites of cv's om aan te geven dat ze kredietwaardig en betrouwbaar zijn. Je krijgt voor burgers en bedrijven al snel een soort wapenwedloop: als iedereen zo'n score heeft en gebruikt, dan ben je verdacht als je niet meedoet.

China, voorbeeldland

Ik denk dat wij onderschatten in hoeverre China een voorbeeldfunctie heeft in de wereld en vooral in Azië. Wat als dit sociale kredietsysteem een succes wordt en leidt tot een optimalisering van de markteconomie? Andere landen zullen volgen en de Chinese staat zou het delen van data met hun systemen verplicht kunnen stellen. We moeten China nu niet gaan zien als een Mordor met een kwaad oog dat vanuit Beijing alles bestuurt. Er is op het gebied van zorg, sociale dienstverlening en onderwijs heel veel

ontwikkeling nodig en het is goed dat de staat hier actief aan werkt. Waarschijnlijk gaan we heel wat deelsucces zien van systemen die onder de vlag van het sociale kredietsysteem aan elkaar gekoppeld worden. Ter illustratie: met *geomapping* kunnen burgers via hun navigatiesystemen *genudged* worden om filevorming tegen te gaan. Of er kan worden voorgesteld dat je gaat carpoolen omdat je dagelijks rond dezelfde tijd naar dezelfde locatie rijdt als drie andere mensen in je flat. Baidu heeft een prototype voorgesteld van eetstokjes die gifstoffen detecteren. Indien op de markt gebracht is dit wellicht een marketinginstrument voor restaurants om vertrouwen in de kwaliteit van het eten te geven. Zo kan precies gemonitord worden in welke stad, straat en restaurant te veel ontoelaatbare stoffen worden gebruikt.

Vergelijking met Nederland

In Nederland verzamelt de overheid ook steeds meer data en hebben we ook krediet-systemen zoals het BKR-register. Deze registraties hebben ook bij ons al sinds jaar en dag gevolgen voor wat je kunt en wat je mag doen. Het is vandaag de dag essentieel om informatie te vergaren om te zorgen dat de markteconomie goed kan functioneren. Als ik mijn huis verkoop wil ik zeker weten dat de makelaar, de koper en de bank betrouwbaar zijn. In China is dat nieuw en omdat de Chinese overheid communistisch is, hebben we al snel argwaan bij zo'n nieuwe ontwikkeling. Wat China bij dit kredietsysteem wel onderscheidt van onze eigen krediet-systemen is de invloed van de overheid. Zo kan China gemakkelijker dan wij platforms dwingen samen te werken en data te delen. En er zijn nauwe banden tussen de Chinese

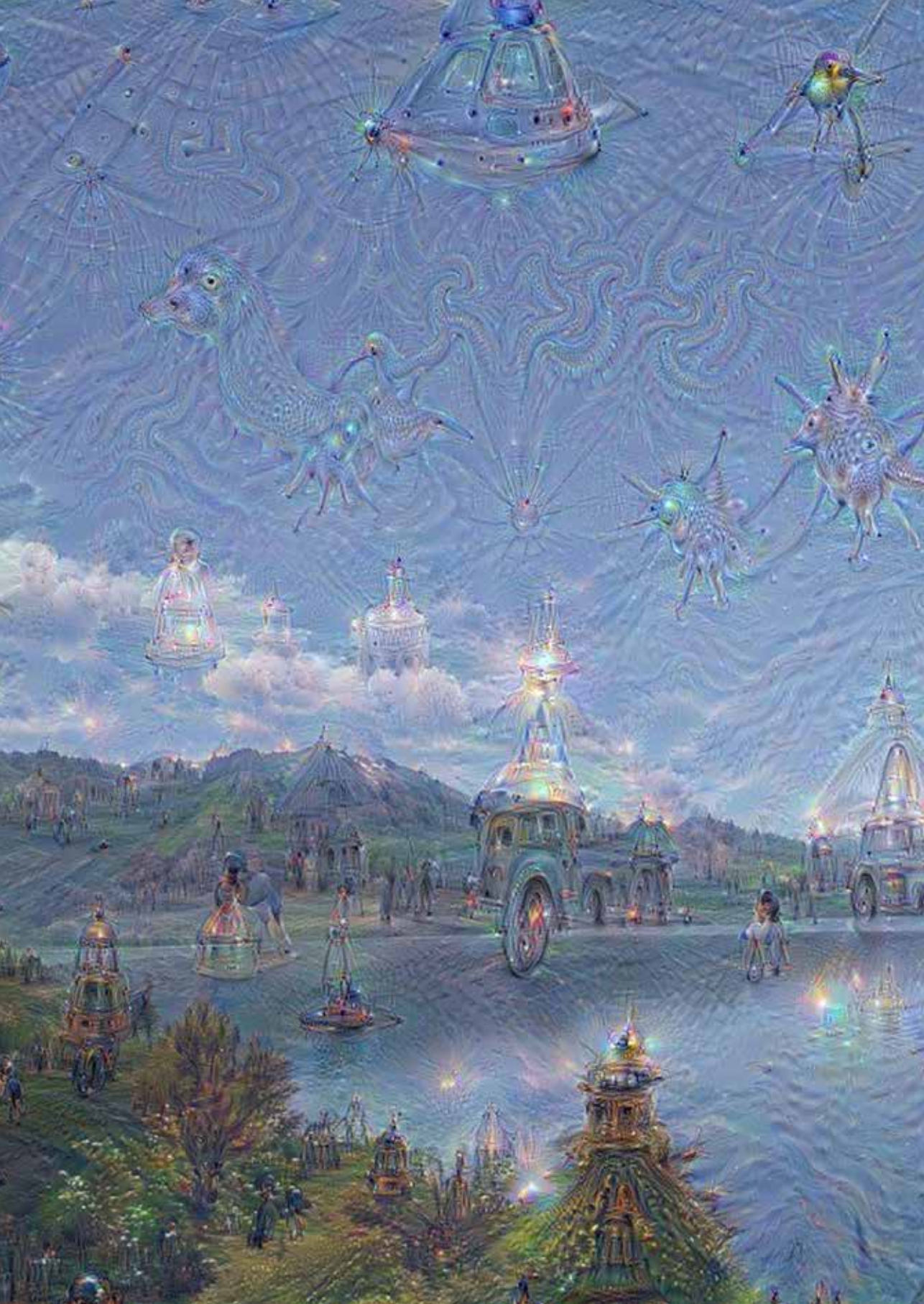
overheid en de grote tech-bedrijven, wat de burgers geen erg sterke positie geeft.

Het leven in cijfers

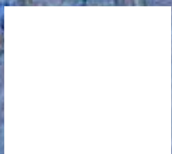
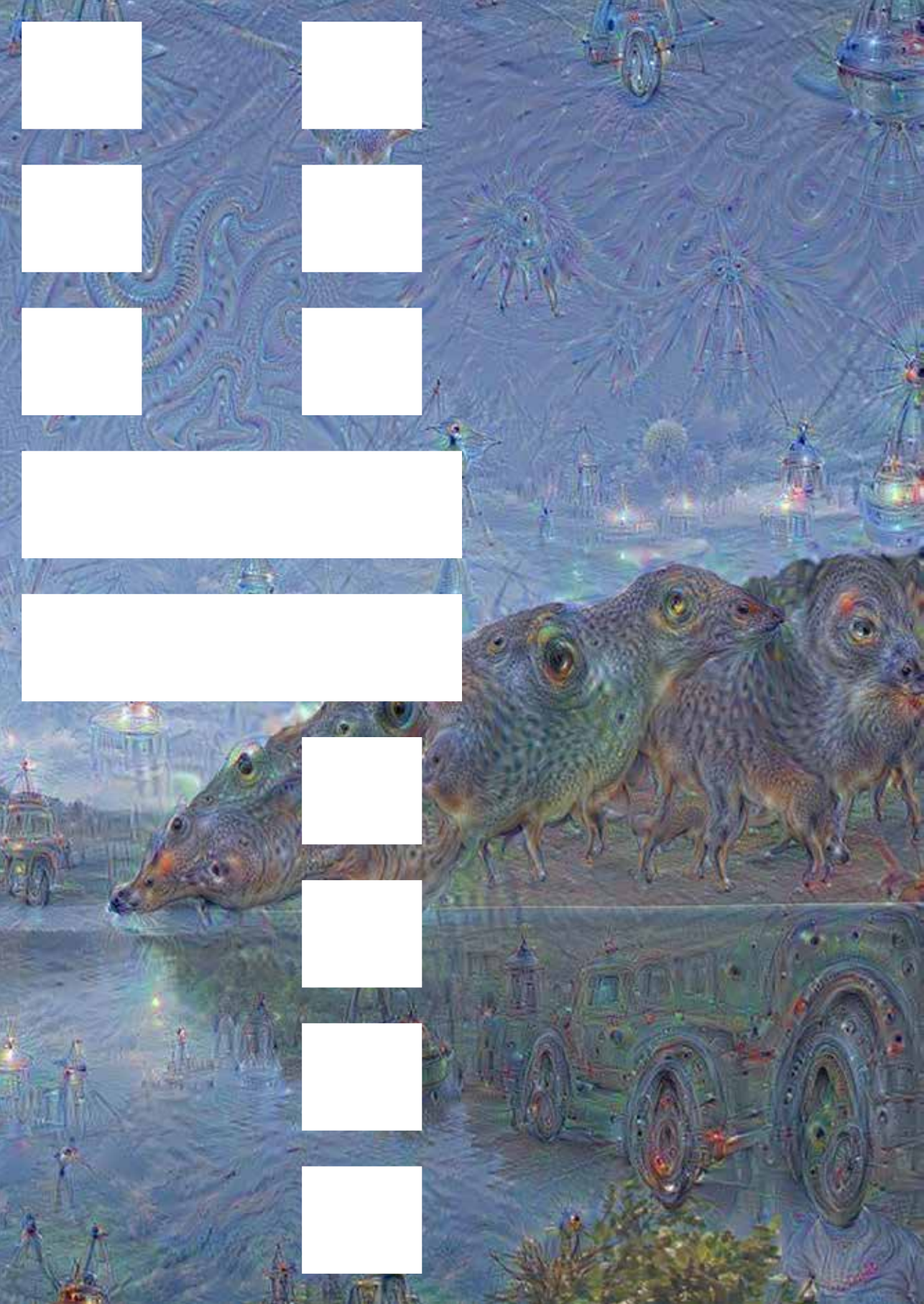
Naast het slagen of falen van zo'n sociaal kredietsysteem in operationele zin moet je je wel afvragen of de ambitie om een burger tot een nummer te reduceren wenselijk is. Nobelprijswinnaar Saul Perlmutter stelt over het huidige academische klimaat dat het "ervoor zorgt dat wetenschappers heel goed zijn in het niet verkwisten van geld en in het niet ontdekken van nieuwe dingen." Als onderzoeker spreekt mij dit aan. Efficiëntie moet niet allesbepalend zijn en

Je krijgt voor burgers en bedrijven al snel een soort wapenwedloop.

we kunnen niet alles in cijfers uitdrukken. Er zijn zaken die in principe onbeheersbaar zijn. Silicon Valley en Peking verschillen daarbij qua jargon niet eens zoveel van elkaar; 'de wereld verbeteren', 'problemen oplossen', 'disruptie', 'nieuwe technologie'. De menselijke conditie wordt veelal gereduceerd tot een probleem dat moet worden opgelost. Zonder te erkennen dat de oplossing in ideale vorm onhaalbaar is en vaak nieuwe problemen veroorzaakt. Big Data bepaalt straks waar we heen rijden en zorgt dat hooligans niet met elkaar knokken. Maar willen we een maatschappij waarbij er geen agressie is omdat het systeem ons zo programmeert of willen we geen knokpartijen omdat we daar zelf voor kiezen? Mijn probleem met dit soort datasystemen is dat ze je verankeren in een bepaalde logica. Ze kunnen je zo in slaap doen dommelen waardoor we niet meer worden gedwongen om na te denken over wat we zelf willen.







4. TIJDSDIMENSIES: WILLEN WE MEER OF MINDER DATA?

Een aanname die fundamenteel is voor de toekomstbeelden die we schetsen is dat data-technologie zich blijft ontwikkelen zoals het tot nu toe gedaan heeft, of dat we zelfs naar een hogere versnelling schakelen. Maar leven we wel toe naar een tijdperk met meer data? Is het voorstelbaar dat de technologie die we ontwikkelen wordt afgeremd, dat we een toekomst met minder technologie tegemoet gaan, of dat het datatijdperk bit voor bit ineen stort? Bij een toekomstverkenning speelt tijd en tijdsperceptie een belangrijke rol. Naast de discussie over technologische trends is het ook belangrijk stil te staan bij de snelheid en richting van verandering. Ontwikkelt datatechnologie zich exponentieel of niet en kijken we naar technologische ontwikkeling vanuit een historisch perspectief of vanuit bepaalde wetmatigheden? Heeft het überhaupt wel zin om je met de toekomst bezig te houden? In dit hoofdstuk gaan we in op verschillende tijdspercepties en presenteren zowel groei- als krimpscenario's in een vier-scenariomodel. Dit is een beproefde methode die onzekerheden in kaart brengt en aan de hand van een assenstelsel mogelijke toekomst voorstelt (Schoemaker 1995). Het framework is bedoeld om een discussie over Big Data en de toekomst te faciliteren.

Hoe kijken we vooruit?

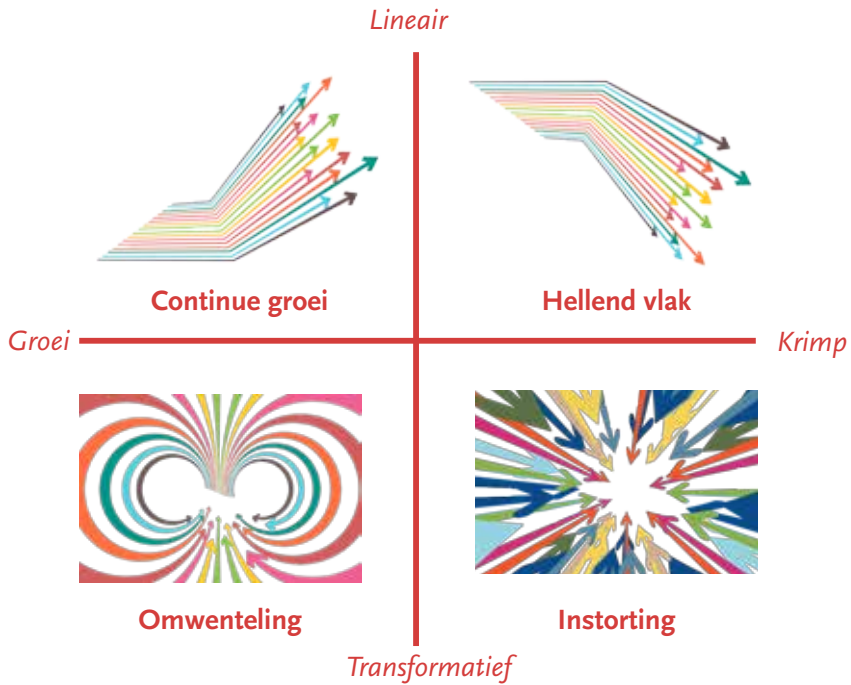
De gerenommeerde toekomstwetenschapper Jim Dator identificeerde tien archetypes die beschrijven hoe mensen naar de toekomst kijken (Bezold 2009). Deze beelden komen allen terug in de Big Data-discussie:

1. *Que sera, sera.* Of het nu God, het universum, determinisme of willekeur is; er is geen patroon.
2. *Oneindige wereld.* Zoals het in den beginne was is het nu ook en zal het altijd zijn.
3. *Ken de cyclus.* Na de winter komt de lente en daarop volgt de zomer; we hoeven alleen maar te weten in welke cyclus we ons bevinden.
4. *De maakbare samenleving.* Handen uit de mouwen, wij bepalen wat er gebeurt.

5. *Viva la Revolución!* We hebben een revolutie nodig om ons te ontwikkelen.
6. *Zondvloed op komst.* We bereiken de grenzen aan de groei en worden geconfronteerd met een toekomst van oorlogen, hongersnoden, interne strijd, gevolgd door donkere dagen.
7. *Stabiliseer.* Stop nu de groei, we moeten pas op de plaats maken, tot een gedecentraliseerd ecologisch evenwicht komen en menselijke stabiele gemeenschappen vormen.
8. *Terug naar de oertijd.* Laten we de industriële samenleving terugdraaien en de natuur omarmen.
9. *Het geestelijk leven.* De toekomst heeft geen werkelijkheid buiten de beelden die we in ons bewustzijn hebben: gebed, meditatie en bewustzijn zijn nodig voor zelfrealisatie.
10. *Na mens komt machine.* Machines nemen ons alle taken uit handen en wij worden overbodig.

De tijdspercepties zijn niet alleen bepalend voor hoe we denken over de toekomst, ze bepalen ook in welke mate het zinvol is over de toekomst na te denken. Als we graag 'que sera, sera' zingen, dan is ons handelingsperspectief beperkt en heeft het niet veel zin om toekomstbeelden en strategieën te ontwikkelen. Als we daarentegen geloven in een maakbare samenleving, waarin onze handelingen richting geven, zijn toekomstbeelden erg belangrijk. Jim Dator presenteerde aan de hand van deze tien beelden een gecondenseerde samenvatting van vier scenario's die goed bruikbaar zijn om de mogelijke ontwikkeling van Big Data te schetsen. De drijvende concepten bij deze paden zijn de groeicurve (lineair of transformatief) en de richting van ontwikkeling (groei of krimp).

1. *Continue groei.* Dit pad gaat uit van het voortbestaan van onze samenleving zoals we die kennen; een samenleving waarin rationalisme, secularisering, wetenschap en techniek, democratisering, individualisering, en voortdurend streven naar vernieuwing, verandering en verbetering centraal staat. De samenleving heeft economische en technische groei als stuwend principe; er is steeds meer behoefte aan data en technologie om deze te verzamelen en te verwerken.
2. *De transformatieve samenleving.* De Nederlandse samenleving zal in de komende jaren fundamenteel veranderen. Er zal een paradigmaverschuiving plaatsvinden die wordt gezien als een 'wedergeboorte'. Voorstanders hiervan zien met de komst van nieuwe technologie zoals Big Data en kunstmatige intelligentie een samenleving met geheel nieuwe waarden ontstaan. Deze verschuiving dringt door in alle domeinen van de samenleving en leidt tot een nieuwe economie, nieuwe zorg, nieuwe overheid, en andere institutionele verschuivingen die leiden tot een complete breuk met het verleden.



3. *Het hellend vlak.* De economische, industriële en technologische groei kan niet voortduren; we zien aan een opeenstapeling van crises nu al dat het tij begint te keren. We moeten leren omgaan met een krimpsituatie en de uitdagingen die gepaard gaan met het managen hiervan.
4. *Technologische instorting.* Gedreven door een tekort aan middelen, gebrek aan voedsel, klimaatverandering, milieurampen, wijdverspreide natuurlijke en door de mens veroorzaakte ziekten, aardbevingen, vulkaanuitbarstingen, buitenaardse invasies, politieke of administratieve onbekwaamheid of een reeks terroristische gebeurtenissen, nucleaire oorlog of een combinatie van al deze factoren glijdt onze huidige beschaving af tot een niveau dat we kenden in de vroege Middeleeuwen – of erger.

Wat opvalt in de discussie over Big Data en de toekomst is dat deze vier ontwikkelpaden niet allemaal gangbaar zijn. *The Big Data Revolution, Singularity is Near, Exponential Organizations*; er is een overvloed aan boektitels met groeiscenario's. Krimpscenario's daarentegen zijn schaars, er zijn vrijwel geen verkenningen of voorspellingen die betogen dat het gebruik van datagedreven technologie zal afnemen of geheel zal worden afgeschaft. Dit sluit aan bij de conclusies uit de *STT Toekomstmonitor*, een onderzoek naar de toekomstperspectieven van meer dan 1000 Nederlanders (Snijders 2016). Uit het onderzoek blijkt dat we in een techno-optimistische tijd leven waarbij meer dan 80% van de respondenten verwacht dat technologie de komende 25 jaar een meer dominante rol in de samenleving zal spelen.

In de onderstaande paragrafen gaan we in op factoren die kunnen leiden tot groei, krimp of transformatie. We beginnen met groeifactoren en beschrijven wetmatigheden die in de Big Data-discussie centraal staan.

Exponentiële groei: drie wetmatigheden

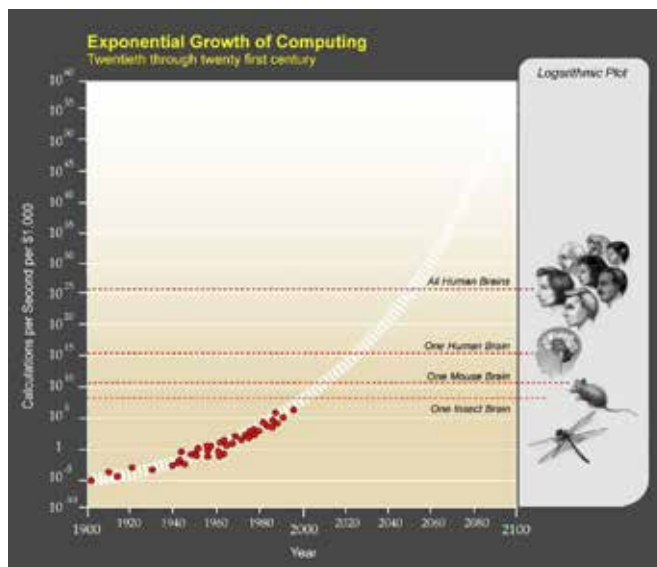
Wiskundigen delen een geheim dat vergaande gevolgen heeft. Ze erkennen namelijk dat sommige problemen te complex zijn om met onze menselijke cognitie op te lossen. De wereld bestaat immers uit eindeloze complexiteit en inconsistente chaotische systemen die zich niet laten voorspellen of berekenen. Een wiskundig voorbeeld dat dit aantoont is het probleem van de reizende verkoper. Stel je voor dat een verkoper dertig steden in Nederland moet aandoen voordat hij terug naar huis keert. Wat is de meest efficiënte route die hij kan nemen? Het vraagstuk is niet alleen van toepassing op het berekenen van transportroutes, maar heeft talloze toepassingen in de praktijk; van het aanleggen van bekabelings- en rioolnetwerken tot het bepalen van de route van een boorkop bij het boren van gaatjes in printplaten. Met alle managers, consultants en brainstormsessies in de wereld kun je deze vraag echter niet oplossen. Er is brute computatiekracht nodig. In 1998 berekenden wiskundigen van de Universiteit van Princeton de oplossing voor een route tussen 15.112 steden in Duitsland. De oplossing kostte 22.6 jaar computertijd op een netwerk van 110 processoren. We hebben niet-menselijke intelligentie nodig om dit soort vraagstukken op te lossen.

Het probleem van de reizende verkoper gaat over exponentiële groei en doet denken aan het verhaal van de uitvinder en de keizer. In opdracht van de keizer bedacht de uitvinder het schaakspel en omdat dit al snel het favoriete spel van de keizer werd mocht de uitvinder een beloning naar keuze bepalen. De uitvinder dacht even na en vroeg of hij één korrel rijst op het eerste speelvakje mocht hebben, twee op het tweede, vier op het derde enzovoorts. De keizer zag het als een nederig verzoek en gaf direct toestemming. De 63 verdubbelingen leiden tot de som van 18.446.744.073.709.551.615 rijstkorrels. Om dit te schenken zou de keizer twee aardbollen – inclusief oceanen – met rijstvelden moeten bebouwen en oogsten. De boodschap van het verhaal is dat exponentiële groei veel sneller gaat dan we intuïtief verwachten. Dit kan onbedoelde gevolgen hebben en men weet niet altijd hoe hij hiermee om moet gaan. Zo kent het verhaal ook verschillende eindes; in het ene wordt de uitvinder aangesteld als speciaal gezant van de keizer, in het andere gaat zijn kop eraf...

Volgens futuroloog Ray Kurzweil kent informatietechnologie eenzelfde exponentiële groei. Het meest bekende voorbeeld van een dergelijke groeipatroon in de IT is de Wet van Moore. In 1965 voorspelde de medeoprichter van Intel dat het aantal transistoren die op een chip passen ongeveer elke 24 maanden zou verdubbelen. Vijftig jaar later gaat deze wet nog steeds op en is dit de voornaamste reden dat een supercomputer nu in je

broekzak past in plaats van in een grote kamer zoals in de jaren '70 van de vorige eeuw.

Zet Moore door? Over deze vraag wordt sinds de initiële voorspelling in 1965 gedebatteerd, en een wijdverspreide conclusie is dat de wet niet eeuwig door kan gaan. "Ik vermoed dat ik Moore's law de komende tien jaar zal zien sterven, maar dat is niet verassend", voorspelde Moore zelf in 2015 (In Courtland 2015). Er zijn fysieke begrenzingen aan de dichtheid waarmee informatie kan worden opgeslagen én er zijn grenzen aan de huidige opslagtechnologie. Moore's Wet heeft in het verleden al verschillende transitieën naar nieuwe technologieën overleefd, dus de tweede beperking hoeft geen spaak in het wiel te steken, maar het ligt niet in de lijn der verwachting dat Moore's Wet ook natuurwetten zal tarten. Over niet al te lange tijd worden transistoren gebouwd die het formaat hebben van slechts enkele atomen; veel kleiner kunnen we niet gaan.



Figuur 4.2 Kurzweil over Moore's wet.

Naast toenemende processorkracht zien we ook dat harddisks steeds meer kunnen opslaan en steeds kleiner worden. Volgens **Kryder's Wet** verdubbelt het aantal bits dat we op een harde schijf kunnen opslaan ongeveer iedere 18 maanden. In 1980 introduceerde fabrikant Seagate de eerste 5.25-inch floppydisk waarop 5MB kon worden opgeslagen en waar een prijskaartje aanhing van \$1500. Nu kun je voor \$600 een harde schijf kopen van 6000GB, ongeveer een miljoen keer zoveel capaciteit voor een veel lagere prijs (vooral als de inflatie ook wordt gecorrigeerd).

Naast Moore's en Kryder's Wet schrijft **Koomey's Wet** voor dat de energie die wordt verbruikt bij computatie iedere 18 maanden halveert. Ook deze trend is opmerkelijk stabiel en is dit al sinds de jaren 50 van de vorige eeuw. Als Koomey's Wet wordt

doorgerekend betekent dit dat de hoeveelheid batterij die nodig is om een bepaalde gecomputeriseerde taak uit te voeren iedere tien jaar met een factor 100 afneemt. Koomey's wet wordt steeds belangrijker, vooral nu energiekosten in de alsmaar uitdijende datawereld een bepalende factor zijn geworden.

Deze wetten, die eerder zelfvullende voorspellingen voor de sector zijn dan werkelijke wetten, schetsen een toekomstbeeld met meer data, meer efficiëntie en meer opslag. Ze zeggen iets over de snelheid van progressie, maar niet over de technologische ontwikkelingen zelf.

Krimp en instorting: een leven zonder data?

Er zijn een aantal factoren die ertoe kunnen leiden dat we minder data gaan produceren. Ten eerste zou – zoals we zojuist lazen – de wet van Moore kunnen ontrafelen. Het kan zijn dat dit ontwikkelpad eindig is en dat we tegen de limiet van de natuur of van ons eigen kunnen aanlopen. Muelhauser en Salamon beschrijven een ander scenario en stellen dat we misschien voor wetenschappelijke problemen komen te staan die we niet voorzien hadden (2012). Het zou goed kunnen dat de ontwikkeling van nieuwe datatechnologie en ons begrip van intelligentie een veld is dat zo complex is, dat het veel meer inspanning vergt dan we hadden verwacht. Misschien is het veld niet te vergelijken met andere wetenschappelijke disciplines en moeten we onze verwachtingen bijstellen (Arbesman 2011).

Naast deze scenario's kunnen externe factoren leiden tot een rem op de ontwikkeling van datagedreven technologie. Er zijn kasten vol met Hollywoodfilms waarin natuurlijke, economische of andere (aardse danwel buitenaardse) rampen leiden tot het einde der tijden of tot een verwoesting van al onze technologie en kennis.

KI-filosoof David Chalmers (2010) denkt echter dat niet externe factoren, maar onze eigen sociaal-politieke ontwikkelingen de meest waarschijnlijke drempel vormen. Wat als we massaal een toekomst met minder data willen? Wat als we dataproductie en kunstmatige intelligentie actief gaan tegenwerken zoals de luddisme-beweging zich in het Engeland van begin 19e eeuw verzette tegen industriële en technologische vooruitgang? Misschien ontstaat er na een aantal catastrofale incidenten met digitale systemen een vorm van neo-luddisme met zoveel macht dat het onze technologie afremt? Maurits Martijn geeft in deze publicatie (p. 91) een doorkijkje naar een dergelijk scenario en schrijft over de dag dat Google valt.

Transformatie: intelligentie explosie

Instorting van onze techno-samenleving kan ook veroorzaakt worden door exponentiële groei van technologie. We kunnen bijvoorbeeld een te grote afhankelijkheid krijgen van datasystemen wat kan leiden tot een 'single point of failure'. Dit houdt in dat we

kwetsbaar worden als we blindelings vertrouwen in een bepaald type technologie en misschien wel ten onder gaan als we geen alternatieve systemen hebben. Toen Google op 16 Augustus 2013 enkele minuten offline ging, crashte het internetverkeer wereldwijd met 40% bijvoorbeeld. Stel dat we ons landbestuur of militaire operaties laten besturen door slimme systemen, maar te maken krijgen met een hevige storing, uitval, of aanval waardoor kerntaken van de samenleving in één klap worden weggevaagd?

Een ander transformatief scenario is het ontstaan van superintelligente systemen die een dominante rol aannemen ten opzichte van de mens. Dit scenario wordt wel eens de intelligentie explosie genoemd. Als mensheid zijn we op verschillende manieren bezig om onze intelligentie te verbeteren. We hacken het brein, we koppelen het aan een computer, of we creëren niet-menselijke intelligentievormen. Intelligentie is, zoals Yudkowsky betoogt, de bron van technologie (Yudkowsky 2008) en als we technologie kunnen gebruiken om intelligentie te verbeteren, dan zou dit kunnen leiden tot een positieve *feedback-loop*. Deze gedachte werd in de jaren '60 van de vorige eeuw al gebezigd door Jack Good, een statisticus en collega van Alan Turing. Het idee is dat als intelligentie ook maar een beetje verbetert, het innovatieproces daardoor versnelt en we zo een exponentieel groeiende vorm van kunstmatige intelligentie zouden kunnen krijgen. Een pure vorm van deze intelligentie-explosie zou een KI-systeem zijn dat z'n eigen broncode herschrijft en blijft herschrijven. Hoe zou zo'n systeem de wereld inrichten? Hoe zouden de objecten om ons heen eruit zien als ze door een dergelijke intelligentievorm zouden zijn ontworpen?

Als we een intelligentie-explosie meemaken, dan zou het één van de belangrijkste gebeurtenissen zijn die ooit heeft plaatsgevonden. Het zou enorme voordelen kunnen hebben: we zouden ongekende wetenschappelijke stappen kunnen zetten, meer ziektes kunnen bestrijden, armoede verhelpen, maar er zijn ook grote potentiële gevaren. Wat als het systeem niet beteugeld kan worden door mensen? Wat als deze intelligentievorm geen sympathie heeft voor mensen?

Nick Bostrom schetst een scenario waarin een generieke KI de simpele opdracht krijgt om zoveel mogelijk paperclips te ontwikkelen (Bostrom 2014). Als het systeem begint met intelligentie die de mens evenaart zou het in eerste instantie misschien paperclips gaan verzamelen, opkopen, of zelf gaan produceren. Het zou daarnaast werken aan het verbeteren van zijn eigen intelligentie, omdat meer intelligentie leidt tot meer manieren om de concurrentie voor te zijn en dus ook tot meer paperclips. Als de KI-paperclipproducent tot de conclusie komt dat mensen het productieproces in de weg zitten zou het systeem mensen kunnen gaan ontslaan, de toegang tot fabrieken kunnen ontzeggen, of zelfs een nucleaire oorlog beginnen tegen de mensheid om maar meer paperclips te maken. Misschien wordt het systeem wel zo 'intelligent' dat het grote gedeeltes van het heelal kan koloniseren en transformeren tot paperclips.

Voor ons klinkt dit als super-domheid in plaats van super-intelligentie. Het KI-systeem begrijpt klaarblijkelijk niet waar het om gaat in de wereld. Het verhaal maakt het echter wel voorstelbaar dat we hele slimme systemen kunnen ontwerpen die geheel niet dezelfde waarden en doelen als wij hebben. Het laat zien dat een vorm van KI die niet specifiek geprogrammeerd wordt om goed te zijn voor de mensheid zeer negatieve gevolgen zou kunnen hebben. De boodschap van Bostrom, die in 2015 herhaald werd door Stephen Hawking, Elon Musk en tientallen KI-experts in een pamflet genaamd *Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter*, is om menselijke waarden mee te nemen in het ontwerpproces van niet-menselijke intelligente systemen. Een belangrijke reden hiervoor is dat het niet aannemelijk is dat systemen deze waarden zelf zullen ontwikkelen.

KI: verwachtingen

De vraag is wanneer KI zich het intelligentieniveau van de mens zal aanmeten. Twintig jaar geleden won het IBM-systeem Deep Blue van een menselijke grootmeester in het spel schaken. Vijf jaar geleden won Watson het spel Jeopardy van menselijke kampioenen. Begin 2017 won Google's AlphaGo via neurale netwerktechnologie en deep learning het spel Go tot drie keer toe van grootmeester Ke Jie, de nummer één in de wereld. Kunstmatige intelligentie is al tijden geen spel meer.

Wel is het bij deze ontwikkeling goed om onderscheid te maken tussen algemene generieke intelligentie die mensen eigen is en de zogenaamde nauwe vorm van KI, van de genoemde systemen. Watson is een zelflerend systeem dat beter dan wie dan ook Jeopardy kan spelen, maar Watson kan niet ganzenborden, laat staan autorijden. Moravec's Paradox speelt hierbij een rol.

Moravec schreef dat wij intelligentie automatisch associëren met iets dat hoogopgeleide mensen moeilijk vinden om te doen, zoals schaken, wiskunde, of een analyse maken van Shakespeare's werk. De paradox is dat het voor digitale systemen relatief eenvoudig is om te schaken of intelligentietesten uit te voeren, maar dat het vrijwel onmogelijk is om net zo bekwaam als peuters te handelen op het gebied van perceptie, mobiliteit en coördinatie.

Het KI-wetenschapsveld staat tot nu toe bekend om haar torenhoge ambities, die ze niet altijd waar kan maken. De bedenker van de term KI, John McCarthy, vertelde in de jaren '60 aan zijn geldschietters bij het Pentagon dat het ongeveer tien jaar zou duren om een machine te maken met een menselijk denkniveau. Herbert Simon van Carnegie Mellon University voorspelde in 1965 dat machines elke vorm van menselijk werk zouden kunnen overnemen. De Amerikaanse marine was nog ambitieuzer en wilde in 1958

op basis van de neurale netwerktheorieën van psycholoog Frank Rosenblatt een ‘denkende machine’ bouwen (Times 1958). De marine presenteerde een ontwerp van een ‘elektronisch embryo’ aan de pers en verwachtte dat deze zou kunnen lopen, spreken, schrijven, voortplanten en zelfs zelfbewustzijn zou tonen. Het project zou 100.000 dollar kosten en binnen een jaar klaar zijn.



De vraag *wanneer* we machines krijgen die zo intelligent zijn als mensen blijft onbeantwoord. Een van de grootste uitdaging hierbij is een filosofische: wat bedoelen we eigenlijk met intelligentie? Neurowetenschappers onderkennen in alle bescheidenheid hoe weinig we nog van het menselijke brein begrijpen. Het gaat bij intelligentie immers om een systeemeigenschap van miljarden cellen die met elkaar verbonden zijn, constant in verandering zijn en input krijgen van en output geven aan het lichaam en de wereld. KI-onderzoeker Luc Steels vergelijkt: “Je kunt een plant uit een tropisch regenwoud proberen na te maken. Dat lukt al nauwelijks, maar stel dat het lukt, dan weet je nog niet hoe die werkt in het geheel van het regenwoud. Er is ook nog de zon, er zijn andere planten, er zijn bijen. Er is het water. Die dragen allemaal bij aan de vorming van de plant. En dat alles verandert voortdurend” (Van Engelen 2015).

Ondanks deze uitdagingen opent het denken over de toekomst voor velen wel een deur naar een KI-systeem dat zich kan meten aan mensen. Uit de *Nationale Toekomstmonitor 2016*, een onderzoek dat STT afnam onder 1000 Nederlanders, bleek dat Nederlanders gemiddeld verwachten dat kunstmatige intelligentie tussen 2040 en 2050 het niveau van menselijke intelligentie voorbij zal streven. Maar hoewel men dit wel verwacht, is men er niet positief over. Ruim 46% van de ondervraagden vindt het een onwenselijke ontwikkeling, 12% is positief en de rest is neutraal of heeft geen mening. De vraag wanneer KI ons voorbij streeft, is in een aantal belangrijke onderzoeken ook aan deskundigen voorgelegd.

Onderzoek 1: 1973

Een van de eerste onderzoeken vond plaats in 1973 en werd uitgevoerd door de befaamde KI-onderzoeker Donald Mitchie. Hij vroeg aan 67 KI-experts hoeveel jaar het zou duren totdat “computers intelligentie demonstreren van een volwassen menselijk

niveau.” Zoals je kunt zien was er destijds geen grote KI-hype en voorspelde de meerderheid dat het pas in 2023 of later plaats zou vinden.

Aantal jaren	Percentage	Jaar
5	0%	1978
10	1%	1983
20	17%	1993
50	19%	2023
>50	25%	>2023

Onderzoek 2: 2006

Moor (2006) hield een enquête op de vijftigste verjaardag van de beroemde KI-conferentie die in 1956 te Dartmouth werd georganiseerd. Opnieuw verwachtte een groot percentage van de respondenten dat het meer dan vijftig jaar zou duren. Ook zien we een aanzienlijke mate van scepsis: maar liefst 41% verwachtte dat machines nooit menselijke intelligentie zou kunnen simuleren.

The earliest that machines will be able to stimulate learning and every other aspect of human intelligence:		
Within 10 years	6	5%
Between 11 and 25 years	3	2%
Between 26 and 50 years	14	11%
More than 50 years	50	41%
Never	50	41%
Totals	123	100%
The earliest we will understand the basic operations (mental steps) of the human brain sufficiently to create machine simulation of human thoughts is:		
Today (we already understand enough)	5	6%
Within the next 10 years	11	12%
Within the next 25 years	9	10%
Within the next 50 years	19	21%
Within the next 100 years or more	26	29%
Never (we will never understand enough)	19	21%
Totals	89	100%
The earliest we will understand the architecture of the brain (how its organizational control is structured) sufficiently to create machine simulation of human thought is:		
Within 10 years	12	11%
Between 11 and 25 years	15	14%
Between 26 and 50 years	24	22%
More than 50 years	44	40%
Never	15	14%
Totals	110	100%

Onderzoek 3: 2016

Een meer recente serieuze poging is uitgevoerd door Müller en Bostrum (2016). Hun onderzoek is uitgevoerd door enquêtes te sturen aan vier groepen met KI-deskundigen. Ongeveer 50% van de ondervraagden verwacht dat KI rond 2040-2050 een hoger intelligentieniveau zal hebben dan wij, en 90% verwacht dit in 2075. De experts verwachten dat KI-systemen binnen 30 jaar na deze datum een vorm van superintelligentie zullen hebben. Dit houdt in dat er een intellect zal ontstaan dat in ieder domein, inclusief de wetenschappen, creativiteit, algemene kennis en sociale vaardigheden, veel slimmer is dan de slimste mens. De respondenten schatten in dat er een kans van 1 op 3 is dat een dergelijke ontwikkeling niet goed uitpakt voor de mensheid.

Conclusies: exponentiële groei

Er is een revolutie gaande! Althans, als we veel van de sprekers, schrijvers en goeroes in het Big Data-domein geloven. Zij beschrijven hoe digitalisering onze productiviteit en tevens ons denkproces uit handen neemt. Ze schetsen een toekomst met een fundamenteel veranderde samenleving waarin auto's, vrachtwagens en vliegtuigen volledig automatisch bestuurd worden, waarin robots zorg- en onderwijstaken overnemen, en waarin de administratieve klasse overbodig wordt. Dit roept vragen op: waar leiden we mensen nog toe op? Waar verdient men geld aan? Hoe verhouden we ons tot systemen die minstens zo intelligent zijn als wij? Misschien gaan we een toekomst tegemoet waarin het leven zo goed en goedkoop geregeld wordt dat niemand zich zorgen hoeft te maken over banen, welvaart, verschillen, en fysiek welzijn. Het zou echter ook kunnen dat we grootse werkloosheid tegemoet gaan, gepaard met sociale en politieke chaos.

Een assumptie achter de meeste toekomstbeelden is dat (data)technologie zich blijft ontwikkelen zoals het tot nu toe gedaan heeft, of dat het zelfs naar een hogere versnelling zal schakelen. Maar leven we wel toe naar een tijdperk met meer data? Dit hoofdstuk geeft een overzicht van een aantal factoren die ertoe kunnen leiden dat technologische ontwikkeling wordt afgeremd of zelfs instort, en gaat in op mogelijke groeiscenario's. Daarnaast gaat het in op tijdspercepties omtrent te toekomstige ontwikkeling van Big Data. In het vooruitblikken op wat Big Data-technologie de komende jaren teweeg gaat brengen is het onmogelijk te voorspellen wat mensen exact zullen doen met de technologie. Het is gemakkelijker om een aantal trends aan te geven die waarschijnlijk zullen doorzetten. Vijftig jaar lang hebben we exponentiële groei gezien in de digitale wereld. Iedere twee jaar is de computerkracht verdubbeld, en het opslagvermogen en de energie-efficiëntie verdubbelden iedere 18 maanden. Dit betekent dat we nu systemen hebben die miljoenen malen krachtiger zijn dan toen ze voor het eerst ontwikkeld werden. Experts zijn het niet eens over hoe lang de wet van Moore – of die van Kryder en Koomey – zullen doorzetten, maar zelfs als de vaart afneemt zullen we

de komende decennia hardware hebben die veel krachtiger is dan we nu hebben. Meer hardware draagt op zichzelf niet bij aan meer intelligentie, maar faciliteert dit proces wel op verschillende manieren. Verbeterde hardware kan er bijvoorbeeld toe leiden dat 'brute kracht'-oplossingen sneller en beter zullen draaien (Muehlhauser en Salamon 2012), hetgeen we nu zien bij de toepassing van gelaagde neurale netwerken en verschillende algoritmes. Goedkopere hardware kan er ook toe leiden dat er meer hardware ontwikkeld zal worden die toegankelijker is voor een breder publiek. Dit kan leiden tot een toename van experimenten en innovatie.

De truc van de duivel

Het begon met de val van Google. Natuurlijk, er waren al langer geruchten dat het niet klopte. Dat het eigenlijk te mooi was om waar te kunnen zijn en wij elkaar maar bleven vertellen dat het waar was. Met de Googleleaks kwam de bevestiging. De leugen was grootser dan zelfs de meest schaamteloze roddelaars hadden durven roddelen. We hadden elkaar jarenlang iets aangepraat dat *niets* bleek te zijn. Al die tijd vroegen we niet door.

We wilden de inconsistenties niet zien, de valse tonen niet horen. Zoals kleine

kinderen vooral niet vragen naar het waarheidsgehalte van de verhaaltjes voor het slapengaan, zo geloofden ook wij dit sprookje net iets té graag.

Google viel hard, keihard. Kort na de onthullingen van klokkenluider 'Z', de 'Edward Snowden van de technologiewereld', die tot de dag van vandaag anoniem is gebleven, was het gedaan. De grootse leugen stond zwart op wit, de oprichters spraken hem in volle overtuiging uit. Op video. Bam. De koers stortte in, de klanten renden weg. Duizenden werknemers namen ontslag zonder op hun vertrekpremie te wachten.

En door de val zagen we pas hoe ver we het hadden laten komen. Van het wereldomspannende luchtkasteel bleek Google slotgracht, ophaalbrug én uitkijktoren ineen. Pogingen van Amerikaans presidente Melania Trump om de minst besmeurde onderdelen van het bedrijf te redden liepen op niets uit. Het was gedaan.

Ja, Google was *too big to fail*, maar de val was zó hard dat het niet meer op te vangen viel. Sector na sector stortte in.

We hadden elkaar jarenlang iets aangepraat dat *niets* bleek te zijn. Al die tijd vroegen we niet door.

De vervlechting van overheid en commercie was helemaal geen synergetisch web – zoals het ons beloofd was – maar bleek een dodelijke strop te zijn voor een ieder die er deel van uitmaakte. Iedereen deed elkaar na en was in essentie in hetzelfde gat gesprongen, Google achterna. Lemmingen, op zoek naar de klif.

Kort erop viel Facebook. Natuurlijk Facebook. Scherpe observatoren hadden dat al wel een beetje kunnen zien aankomen. In 2016 bleken er fundamentele fouten te zitten in de *metrics* van Mark Zuckerbergs bedrijf. Facebook gaf zijn klanten jarenlang click- en kijkcijfers door

die niet klopten. Jouw video is tienduizend keer bekeken. Jouw artikel is een miljoen keer gedeeld. *Niet dus*. Als er toen één slimmerik zou zijn geweest die dit écht zou hebben doorgerekend, dan zouden we toen hebben ingezien hoe onnatuurlijk groot het percentage gebakken lucht was. Dat zou misschien nog op tijd zijn geweest. Al snel volgden overheden en overheidsinstanties. Iets trager, uiteraard, maar toen zij eenmaal vielen, was het Domino D-Day. Die paar criticasters die er al jaren op tamboereerden dat het overnemen van dit evangelie het domste was wat overheden konden doen, kregen met terugwerkende kracht gelijk. *I told you so*.

Natuurlijk werden de geheime diensten, de terrorismebestrijders en de opsporingsinstanties het hardst geraakt – zij waren immers *all in* gegaan. Maar het bleef niet beperkt tot het veiligheidsdomein. Bij de meeste belastingdiensten was het *nudging*-gebabbel ook al jaren verheven tot de taal die gesproken diende

te worden. Net als in het sociale domein, in de kindercare, de fraudebestrijding, de werkloosheidsindustrie.

Data. Overall draaide het om data. De ene keer met het woord ‘Big’ ervoor, de andere keer gevolgd door de twee lettergrepen *mi-ning*. In de loop der tijd kwamen er woorden bij. Deep learning. Machine learning. Artificial Intelligence. Datagestuurde besluitvorming. Algoritmisch [vul maar in]. Door de val van Google zagen we ineens dat al deze woorden – en daarmee de miljardenindustrieën, de bestsellers en de academische substromingen die eromheen waren ontstaan – precies dat waren: woorden. Woorden, die in essentie op hetzelfde neerkwamen.

Als de grootste truc van de duivel was dat hij ons ervan wist te overtuigen dat hij bestond, dan is toch zeker zijn een na beste truc dat hij ons zo lang heeft doen geloven dat hij erin geslaagd was ons tot datasubjecten te reduceren.

ICT: Onomkeerbaar, weest gewaarschuwd

In plaats van een eenzijdige focus op Big Data gaat dit artikel in op de ontwikkeling van allerlei toepassingen waarin ICT een belangrijke rol speelt. De analysemethodes die voor Big Data gebruikt worden komen immers in verschillende gedaantes in vele toepassingen voor, bijvoorbeeld bij kunstmatige intelligentie of in combinatie met andere kennisgebieden zoals nanotechnologie. We hebben te maken met een verschijnsel met vele gezichten.

Optimisme

We staan aan het begin van een vierde technologische revolutie die onze manier van leven ingrijpend zal veranderen (Schwab 2017). Informatie- en communicatietechnologie spelen daar een hoofdrol in. Veranderingen nemen in een hoog tempo toe. Tekenend is een recente uitspraak van voormalig president Obama: “Tijdens mijn presidentschap veranderde de zelfrijdende auto van scifi in opkomende realiteit” (*Pittsburgh Post-Gazette*, 19-09-2016). Het World Economic Forum jubelde in 2014: “Today we are seeing a data boom rivaling the Texas oil boom of the 20th century and the San Francisco gold rush of the 1800s. It has spawned an entire support industry and has attracted a great deal of business press in recent

years.” (Billao-Osorio *et al.* 2014). Dit beeld trekt het WEF in 2016 door, hoewel ze kanttekeningen plaatsen bij de verstoringende rol van politiek (Baller *et al.* 2016). Het tempo van de veranderingen versnelt nog steeds en wordt aangeduid als exponentiële groei, ofwel technologieën die sneller groeien dan het tempo van Moore’s Law zoals kunstmatige intelligentie, robotisering, quantum computing, en cybersecurity. Doordat kennisgebieden samenkomen treft dit steeds meer de mens in heel zijn wezen. Het Rathenau Instituut is hierover duidelijk: “Mens en technologie versmelten zo snel dat we kunnen spreken van een intiem-technische revolutie” (van Est *et al.* 2014).

Waarschuwingen en pessimisme

Een waarschuwing uit hetzelfde Rathenau-rapport “overheden, bedrijven en andere burgers [kunnen] intieme technologieën inzetten om invloed, of zelfs dwang, op ons uit te oefenen.” In hetzelfde jaar (2014) komt Stephen Hawking met een verdergaande doemgedachte en zegt: “the development of full artificial intelligence could spell the end of the human race.” Volgens hem zijn we op de korte termijn geïnteresseerd in *wie* de ontwikkelingen in de hand kan houden, maar op

de langere termijn is de vraag eerder òf de ontwikkelingen in kunstmatige intelligentie überhaupt beheersbaar zijn.” Hij wordt in deze visie gesteund door een aantal toonaangevende wetenschappers en publieke figuren als Elon Musk. De waarschuwingen zijn niet nieuw.

Norbert Wiener, de grondlegger van de cybernetica, voorzag in 1962 al dat de 'automation' weliswaar effectief kon zijn, maar ook gevaarlijk en onbeheersbaar (Wiener 1999). In Frankrijk meldden twee briljante *'fonctionnaires'* Nora en Minc [1978] “de moeilijke alchemie van de samenleving (l'État) en 'informatique'”. Wie de studie leest kan alleen maar diepe bewondering hebben voor de trefzekerheid ervan; ICT werd als politieke issue neergezet. In talrijke studies daarna, overstemden het techno-optimisme meestal de maatschappijkritische studies. Daarin lijkt momenteel een kentering te komen.

Intermezzo: Big Data

Big Data is het magische toverwoord dat vaak zonder precieze duiding terugkomt, maar dat wijst op zowel onbegrensde mogelijkheden als op bedreigingen. Juist die bedreigingen krijgen steeds meer aandacht: grove aantastingen van de privacy van burgers bijvoorbeeld, zoals recentelijk aannemelijk werd gemaakt in een drie jaar durend journalistiek onderzoeksproject van de Correspondent. Generaal Michael Hayden, voormalige CIA-directeur, zei “We kill people based on meta data” (Martijn en Tokmetzis 2016). Anekdotisch is een experiment, door Mutchler, PhD-student in Stanford: hij verzamelde gedurende een

maand metadata van smartphones. Op basis daarvan kon hij talloze privé-gegevens achterhalen, zoals een compleet medisch dossier en aan welke wapens iemand de voorkeur gaf. *Big Data, do it yourself.*

We kunnen alleen constateren dat we ons aan het begin van de vierde technologische revolutie bevinden.

Valt er iets over de toekomst te zeggen?

We kunnen alleen constateren dat we ons aan het begin van de vierde technologische revolutie bevinden. De richting ervan valt moeilijk te voorspellen, de pogingen daartoe laten heel verschillende beelden zien. Afgezien van eventuele economische en sociale disrupties zijn ten minste vijf, onderling afhankelijke, factoren te onderscheiden die van invloed kunnen zijn op die richting.

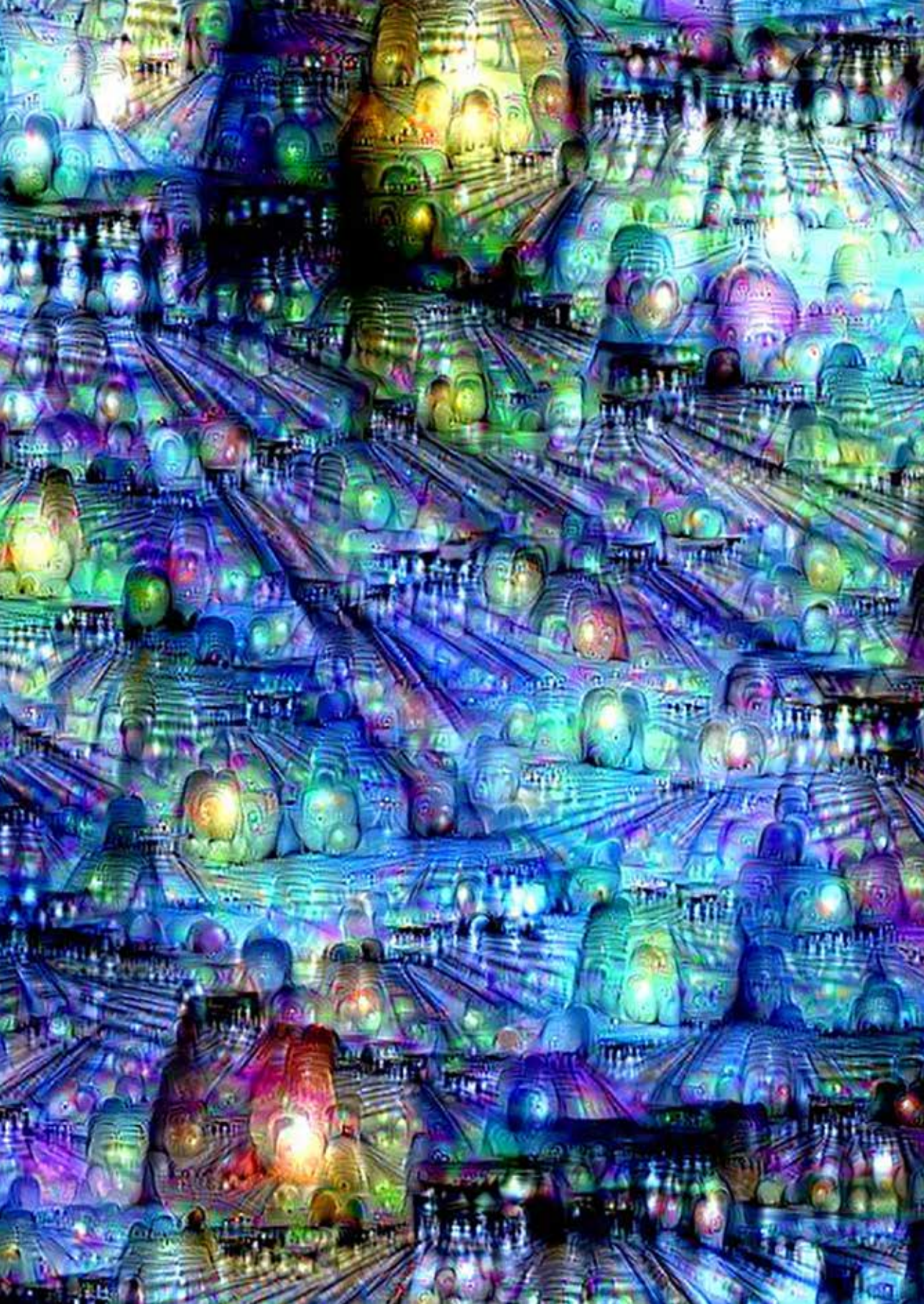
1. Het risico van een aanval op de ICT-infrastructuur. Deze kan systemen volledig verlammen of vernietigen, denk aan de dDos-aanval op banken in 2013. Aanvallen kunnen bovendien worden uitgevoerd door 'lone wolves'.
2. Omvang en niveau van het personeel betrokken bij innovatie, ontwikkeling en onderhoud van systemen. Personen met een bèta-achtergrond zijn schaars: in Europa wordt in 2020 een tekort voorzien van 900.000 personen met ICT-gerelateerde kennis, voor Nederland 56.000. Thans heeft slechts een op de tien afstudeerders een bèta-achtergrond.
3. Disfunctioneren van de systemen zelf. Bij de rijksoverheid is dit bekend, ook in andere sectoren komt het veel voor. Verstoringen komen vaker voor

naarmate systemen complexer worden. Belangrijk is of systemen voldoende op hun werking getest (kunnen) worden. Personeelstekorten doen het ergste vrezen.

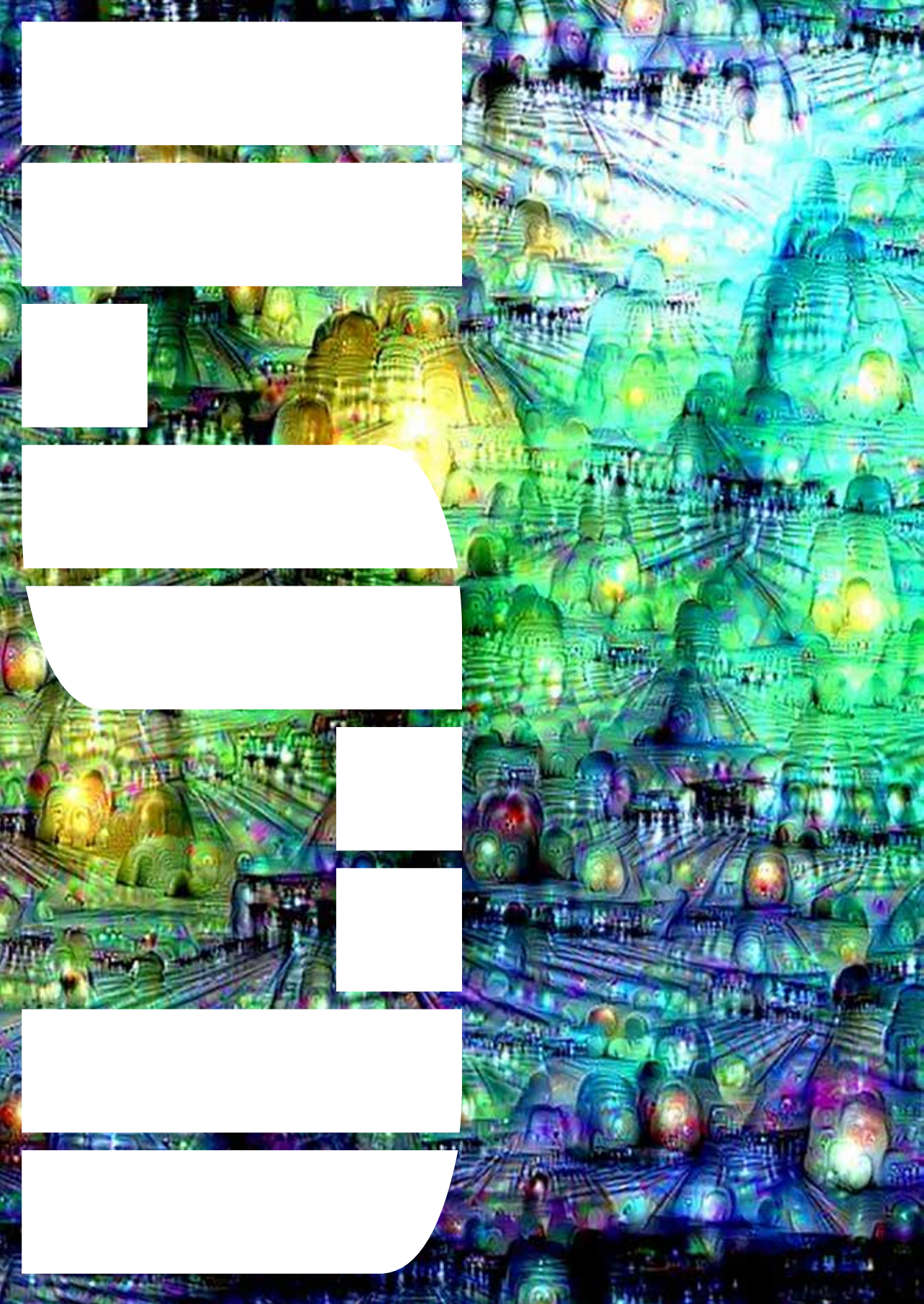
4. De druk op de tijdspanne tussen innovatie en markt. Te korte trajecten leiden tot slecht geteste systemen. Een voorbeeld is de Galaxy Note 7 van Samsung.
5. Het vertrouwen van de gebruiker. Gebrek hieraan haalt de digitale economie onderuit. Steeds meer nadruk zal daarom moeten liggen op ethiek en vertrouwen. Inbreuken op bijvoorbeeld privacy zullen steeds minder worden geaccepteerd, zeker nu bij het brede publiek steeds duidelijker wordt dat gedrag op verschillende niveaus met grote nauwkeurigheid is te voorspellen.

En nu?

De vraag is of Stephen Hawking gelijk heeft. Ik aarzel. Maar veel belangrijker is of dergelijke voorspellingen überhaupt wel mogelijk zijn. Los van sociale en economische disrupties impliceren bovengenoemde factoren onzekerheden waarvoor nauwelijks effectieve (bestuurs)structuren bestaan. Bovendien heeft slechts een op de tien politici een bèta-achtergrond... Het verdient daarom overweging om een gezaghebbend en breed opgezette techno-adviesraad in te stellen die jaarlijks een vinger aan de pols kan houden. Essentieel is dat dit orgaan doorzettingsmacht achter de hand krijgt.







5. NIEUWE VERHALEN

De toekomst wordt gevormd door toekomstbeelden die men vandaag heeft. Vandaar dat het van belang is om bestaande toekomstbeelden te onderzoeken en actief te werken aan nieuwe toekomstbeelden. Dit hoofdstuk bouwt voort op het vorige hoofdstuk en presenteert verhalende groeiscenario's. De scenario's proberen niet zozeer om de toekomst te voorspellen, maar ze proberen een aantal belangrijke beweegrichtingen te identificeren die bepalend zijn voor onze toekomst (Giaoutzi en Sapiro 2013). Het is niet de bedoeling dit hoofdstuk weg te leggen om over twintig jaar te toetsen of de toekomstbeelden kloppen, maar juist om ze nu – als beleidsmakers, strategen en toekomstverkenners – te gebruiken als kapstok voor een discussie over de toekomst van data en de samenleving. Dit geldt zelfs als ze een tikkeltje onvoorstelbaar zijn:

“Say almost anything bold about the future and you will almost certainly sound ridiculous to someone, probably including most people in the future. That’s fine. The future should be our theater. It should be fun and wild, and force us to see everything in our present world anew” (Lanier, 2013).

In dit hoofdstuk bespreken we scenario's die het resultaat zijn van interdisciplinaire workshops (zie Bijlage: Aanpak). We beginnen met vier verhalende scenario's waarin personages ons door hun toekomstige datagedreven levens leiden. Deze scenario's gaan in op de machtsdimensie van Big Data: wie is de eigenaar van data en hoe regelen we het beheer ervan?



Scenario 1: Life as a Service

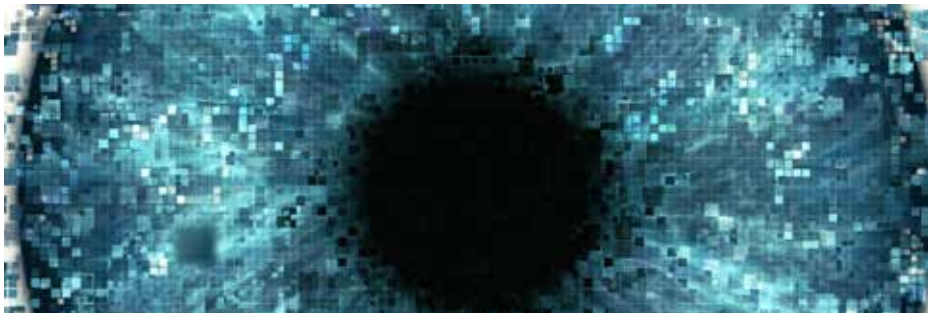
Het is zaterdagochtend en Mascha haalt de boodschappen uit de tuin. “Verse croissants... een smoothie... ahh, was het wasmiddel op?”, mompelt ze. Op basis van haar aankoopprofiel heeft het *Predictive Delivery Systeem* inkopen gedaan en met drones naar haar huis gebracht. “Graag nog wat olijfolie en mozzarella voor die brunch met Sam morgen” zegt ze. Een belletje rinkelt en de producten verschijnen op de boodschappen- en takenlijst op de muur. “Oh ja, en doe maar een vestje erbij, Jean-Paul”. Mascha’s kleding wordt door een online modestyliste uitgekozen, die haar wensen door en door kent. Wekelijks wordt haar garderobe geactualiseerd en grotendeels vervangen. Dit gebeurt aan de hand van Mascha’s agenda, haar reisprofiel, het klimaat, de contacten die ze zal hebben en ook haar meubels, die eveneens regelmatig worden vervangen.

Tijdens het ontbijt bekijkt Mascha haar carrièreprofiel. Ze is al zeven jaar in dienst bij een grote internationale organisatie als informatie-architect, heeft een aanzienlijk netwerk en haar diensten worden erg gewaardeerd. Tegelijkertijd komt ze maar langzaam hogerop in de organisatie, vertoont haar stemmingsmeter dipjes en begint haar frisheidsscore af te nemen. De data spreekt: ze is toe aan iets nieuws. Promotie zou mooi zijn, een carrièreswitch, een sabbatical, of een hele nieuwe uitdaging. Alphagamma, de wereldwijde dataleider die ook haar dataprofiel bijhoudt, adviseert haar om Life as a Service te activeren. Ze moet hiervoor haar profiel opwaarderen, maar kan dit kosteloos doen door haar data voor een jaar vast te zetten op ‘ongelimiteerd delen’. Mascha zet Life as a Service aan en gaat in gesprek met Kit, een virtuele assistent. Kit adviseert Mascha op basis van een Natural Thought Processing Assessment (NTP-Assesment) om in gesprek te gaan met haar werkgever en te vragen om overplaatsing naar de afdeling strategie, waar een plek vrijkomt. Kit heeft al een voorbeeldbericht opgesteld aan Suzan, haar directe manager. Het bericht is slim bedacht en precies afgestemd op de voorkeuren van Suzan. Zo weet het systeem dat Suzan het liefst korte, persoonlijke berichten ontvangt die snel ter zake komen. Tegelijk presenteert Kit statistieken over wanneer Suzan haar mail leest, hoe ze zich verhoudt tot Mascha en op welk moment ze het meest geneigd is positief te reageren op geldgerelateerde zaken. Kit weet dat het dataprofiel van Suzan ook beschikbaar is en biedt tegen een eenmalige betaling een virtueel gesprek

met haar aan. Het bedrag is te hoog voor Mascha's budget, maar Kit verzekert haar dat haar kans op promotie met 71% zal toenemen als ze deze dienst afneemt. Ook weet hij Mascha gerust te stellen over haar financiële situatie en stelt dat haar stedentrip deze maand nog zonder betaling kan worden geannuleerd.

Mascha gaat ervoor. Vijf minuten later zit haar baas bij haar aan de keukentafel. "Hiiii, ik zag je bericht en dacht, ik schakel meteen even." Het is net echt. Op basis van bestaande beelden, gespreksdata, en de NTP-assesment is een feilloze holo-constructie van Suzan nagemaakt. Ze lijkt op Suzan, spreekt als Suzan, en beweegt als Suzan. Spannend!

Kit adviseert Mascha telkens wat ze het beste kan zeggen en geeft aan hoe ze overkomt op Suzan. Uiteindelijk wordt het gesprek vier keer geoefend en telkens gaat het gesprek net iets beter. Mascha is zelfverzekerd geworden en twijfelt niet meer aan haar eigen kunnen. Ze laat Kit het sollicitatiebericht verzenden, leest de disclaimer van Alphagamma over de onvoorspelbaarheid van de werkelijkheid, en wacht af. Die middag nog is er een holo-oproep en jawel, het is haar baas. "Hiiii, ik zag je bericht en dacht, ik schakel meteen even." Het is de eerste keer vandaag dat ze een echt mens spreekt.



Scenario 2: Brave nieuwe wereld

Paul heeft een afspraakje. Hij is aan de late kant en zet zijn OV-auto op handmatig. Hij rijdt over de snelweg terwijl z'n persoonlijke digitale assistent nog een laatste keer doorneemt met wie hij uit eten gaat. Sanna heet ze. Ze is net afgestudeerd, werkt nu als data-ingenieur op een basisschool en houdt van infomusic. Pauls stuur trilt. Hij wordt toegesproken door een spreekbot van de verkeerscentrale. "Op basis van uw rijgedrag voorspelt ons datasysteem Prophet een potentieel kritieke verkeerssituatie. Uw automatische piloot is ingeschakeld en kan tot de aankomst op uw bestemming niet worden gedeactiveerd." Misschien maakt dit afspraakje Paul nerveuzer dan hij denkt? Data over zijn emoties, zijn hartslag, lichaamstemperatuur en rijgedrag is door de centrale verkeersregie bestempeld als risicovol.

Sinds Prophet is ingevoerd is het aantal verkeersongelukken geminimaliseerd. Het laatste dodelijke verkeersongeval is al van jaren terug. Onvoorstelbaar dat er in 2015 alleen in Nederland al 621 verkeersdoden waren! Big Data heeft de samenleving stukken veiliger gemaakt. Niet alleen verkeersongelukken, maar ook het aantal aanslagen, moorden, vernielingen, en vechtpartijen is sterk afgenomen. Er zijn veel minder spanningen tussen bevolkingsgroepen en, nadat data uitwees dat mensen in relaties een gezonder en meer maatschappelijk geëngageerd leven leiden, wordt actief gestuurd op stabiele liefdesverhoudingen. Micro-agressies houden de gemoederen wel bezig. Paul luistert op het journaal naar een schokkend verhaal over een oudere dame die verantwoordigd reageerde toen haar buurman geen Nederlands sprak. Zelfs na een digitale berisping en een direct geïncasseerde boete stak de vrouw haar ontstemdheid niet onder stoelen of banken. Ze wordt morgen op basis van gepersonaliseerde wetgeving voorgelid en berecht.

Paul wordt afgezet voor het restaurant waar hij zijn moet, terwijl zijn voertuig zichzelf parkeert en weer oplaadt. Sanna is er nog niet en Paul loopt een rondje over het plein. Er zijn erg veel mensen op de been voor een woensdagavond, maar *automatic crowd control* (ACC) doet zijn werk. Door middel van gecontroleerde lampen en gepersonaliseerde berichten zorgt het systeem ervoor dat men geen hinder heeft van elkaar en dat er geen opstootjes kunnen plaatsvinden. Paul steekt over en hoort iemand zijn naam roepen. “Niet treuzelen hier”, fluistert het ACC-systeem in zijn oor als hij omkijkt, maar hij negeert de boodschap en loopt op het meisje af. “Ha Sanna?!” “Hi Paul!” Ze ziet er in het echt nog leuker uit dan virtueel. Ze lachen naar elkaar, het is liefde op het eerste gezicht. Paul is 22 jaar en heeft een gelukkig leven voor de boeg. Het rijkspartnersysteem scoort weer.



Scenario 3: Een prettig profiel

Sharon en Lani zijn 15 jaar getrouwd en financieel gaat het voor de wind. Het geld stroomt binnen via betalingen die zij vooral krijgen voor hun data. Toen Sharon Lani ontmoette studeerde ze nog sociale geografie. Hoewel ze qua cijfers gemiddeld niet bijzonder hoog scoorde ontving ze de faculteitsprijs voor beste masterthesis. Een mooie

prestatie, vooral naast haar bestuursfunctie bij de studievereniging. Voor de prijs kreeg ze een geldbedrag dat vrij te besteden was, maar voor haar dataprofiel ‘semester 2, 2018’ ontvangt ze nog steeds royalty’s. Sharon was al vroeg gebiologeerd door de quantified self-beweging en legde tijdens haar studietijd met fitbits en andere trackers vast wanneer ze sliep, hoeveel ze wandelde, wat ze at, en waar ze was. Haar profiel was rijkelijk gevuld – vooral voor een profiel uit die tijd – en is daardoor nu nog altijd in trek bij EduTech-bedrijven. Maandelijks gaat Sharon’s gekoppelde dataprofiel – een virtuele verstrengeling van haar dieet, studiecijfers, liefdesleven, en veel meer – honderden keren over de digitale toonbank. Het wordt gewogen met profielen van andere succesvolle studenten om zo tot nieuwe inzichten en voorschriften te komen. Het mooie is dat ze iedere keer dat haar data gebruikt wordt een microbetaling krijgt. En dat niet alleen voor ‘semester 2, 2018’, maar ook voor de talloze andere datafasen uit haar leven die ze beschikbaar heeft gesteld. De andere semesters natuurlijk, maar ook haar vrijwilligersprofiel van de tijd dat ze bij de voedselbank werkte, de dataverzameling over protestmarsen die ze in haar leven bezocht, haar vergaderdata en tandenpoetsgewoontes. Dataprofielen zijn er niet alleen op individueel vlak. Als organisatie, vriendengroep of voetbalteam kun je een gezamenlijk profiel opstellen om zo geld te genereren. Zo ontvangt Sharon samen met Lani relatieroyalty’s voor hun partnerprofiel.

De betalingen komen binnen via persoonlijke datakluisen, en inmiddels heeft 94% van de bevolking één of meerdere kluisen afgesloten. De datakluisen werken via blockchaintechnologie, zonder tussenkomst van banken of overheden. In de kluis staat exact welke data er over je worden opgeslagen en wat je datavoorkeuren zijn. Wil je dat je transportdata gedeeld worden met je verzekeringsmaatschappij? Mag de supermarkt jouw aankoopgedrag opslaan om persoonlijke advertenties te sturen? Het bijhouden van alle persoonlijke datahandel en het creëren van nieuwe data-ervaringen is veel werk. Sharon en Lani hebben daarom IndiviData ingezet, een persoonlijke financiële assistent die adviseert hoe je meer geld uit je leven kunt halen. Deze software maakt, gebaseerd op een intakegesprek en regelmatige evaluatiegesprekken, honderden datakeuzes per dag. Als je over straat loopt onderhandelt het systeem met alle digitale actoren om je heen en bepaalt in flitstransacties welke reclames je te zien krijgt, welke data worden opgeslagen om je profiel te verrijken, en welke data je met wie deelt. Alles gaat automatisch, een betaling is geen fysieke transactie meer. IndiviData geeft daarnaast advies in welke dataprofielen jij het beste kunt investeren en hoe je je eigen profiel het beste in de markt kunt zetten. Je data worden immers meer waard als anderen ervan weten, als ze je volgen en als je erkenning krijgt.

Sharon en Lani zijn blij met de inkomsten en kunnen er prima van rondkomen. Hun gezamenlijke profiel wordt met de dag meer waard en levert een leuk zakcentje op. Vanuit een financieel oogpunt zou scheiden, volgens IndiviData, een slechte zet zijn.



Scenario 4: Ubuntu

Jin is een nieuwkomer bij AI4All en werkt aan gratis KI-diensten voor de gemeenschap. Hij werkt aan het ontwikkelen van software die mensen extra intelligentie biedt; bijvoorbeeld om te kunnen spreken met personen die je taal niet kennen, om je geheugen uit te breiden, om je rekenkundig vermogen te versnellen, et cetera. Hij zit een filmpje te kijken van een van de oprichters van het bedrijf. “Hoe kan het zo zijn dat iedereen geld verdient aan data, behalve diegene die de data maakt? We leven in een wereld waarin informatie belangrijker is geworden dan fysieke arbeid, maar toch wordt er van ons verwacht dat we onze informatie zo maar opgeven en overhandigen aan overheden en marktpartijen die het omzetten in hun eigen kapitaal.” AI4All ontstond uit de invloedrijke protestbeweging Ubuntu, die zich tegen intelligentiemisbruik door grote bedrijven en overheden verzette. Ubuntu is van oorsprong een Afrikaanse filosofie die zoiets betekent als ‘Ik ben omdat wij zijn’. Met slogans als ‘free my intel’, ‘wisdom for the crowds’, en ‘ik ben wij’ kreeg Ubuntu veel volgers. Ze organiseerden data-stakingen, waarbij persoonlijke data werden gemanipuleerd om centrale systemen van de wijs te brengen, en vormden al snel een politiek platform. De Ubuntu-gemeenschap bestaat inmiddels uit duizenden initiatieven zoals ziekenhuizen, universiteiten, boerderijen, en lokale softwarehubs. Binnen de gemeenschap zijn alle diensten kosteloos af te nemen voor personen die zich voldoende inzetten voor de gemeenschap. Zo werkt Jin voltijd bij AI4All en kan zo van alle Ubuntu-diensten gebruik maken. Naast voltijdscolllega’s zijn er honderden anderen die een paar uurtjes per week aan AI4All-projecten sleutelen, naast hun andere functies binnen Ubuntu. Mensen die niet werken voor de gemeenschap, maar voor geld, betalen om van de diensten gebruik te maken.

De beweging is decentraal georganiseerd, maar internationaal verbonden want ‘Big Data kent geen grenzen’. Data-vluchtelingen uit landen zoals China en de Verenigde Staten zijn dan ook welkom in de beweging en AI4All steunt hun toevlucht naar Nederland actief door ze kansen aan te bieden in hun Ubuntu-organisatie. Jin is zelf Chinees en heeft met zijn eigen ogen ervaren hoe kunstmatige intelligentie tot ongelijkheid leidt als er ongelijke toegang toe is. Gelukkig heeft hij nu een veilige thuishaven

in Nederland gevonden en is er voor Jin nog zat werk te doen. Een handjevol Chinese miljardairs heeft daar in korte tijd de beurs ontregeld door geavanceerde AI-systemen voor hun karretje te spannen. Dit leidde er niet alleen toe dat armen veel armer werden, maar ook dat de gewone Chinees politieke invloed en vrijheden verloor. Gelukkig kon Jin als datavluchteling naar Nederland afreizen en is er nog zat werk te doen. Na een lange dag werken zucht hij “blockchain... 4D-printers... neurale netwerken... 哎呀, wat loopt men hier achter!”

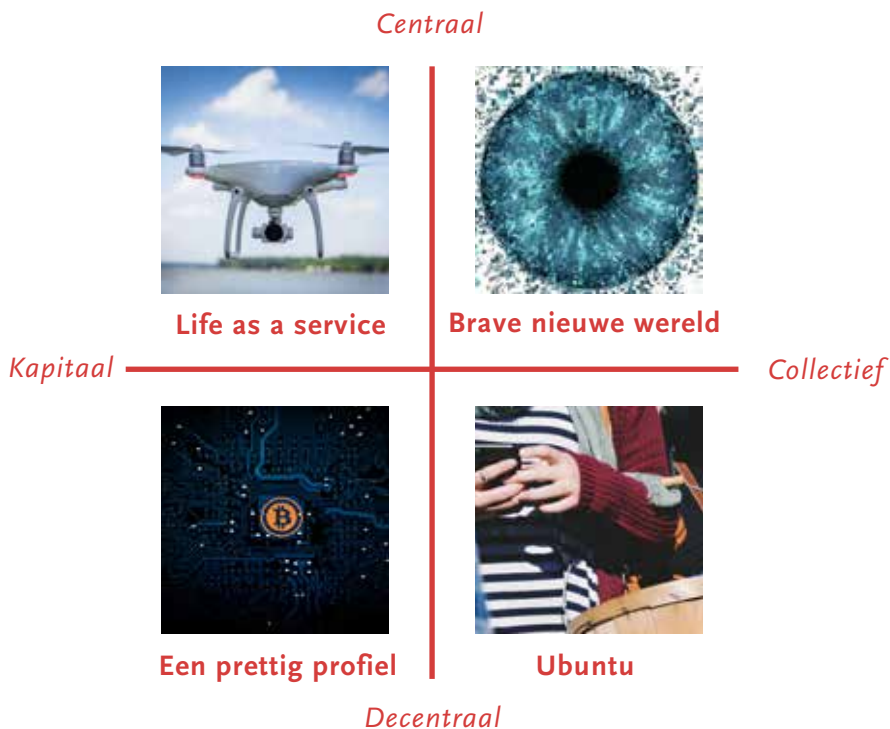
Analyse: Big Data is politiek

Een discussie over de toekomst van Big Data wordt vroeg of laat een politieke discussie. Het onderwerp is misschien technologisch van naam en tot op bepaalde hoogte ook van aard, maar als het gaat om het gebruik van data die een belangrijke rol spelen in ons leven en onze samenleving moeten ook politieke vraagstukken beantwoord worden. Van wie zijn de data? Hoe beschermen we onze data? Hoe voorkomen we dat data misbruikt worden om mensen te manipuleren? Waar ‘kennis is macht’ in de zestiende eeuw al een geveugelde uitspraak was, kan deze in het Big Data-tijdperk worden geüpdate tot de uitspraak ‘data is macht’. In de datadiscussie kunnen belangrijke spanningsvelden worden gevat in de volgende vraagstukken:

- 1. Waarom verzamelen we data? Doen we dit vanuit winstoogmerk of voor de gemeenschap?**
- 2. Hoe organiseren we onze data, doen we dit centraal of decentraal?**

Geïnspireerd door Kostakis en Bauwens’ studie naar peer2peer-platforms (2014) zijn de hiervoor beschreven machtsscenario’s ontwikkeld in de onderstaande matrix. We zullen de scenario’s hieronder één voor één analyseren en afsluiten met een paragraaf waarin we de vier scenario’s met elkaar vergelijken.

- Scenario 1: Life as a service. Big Data wordt gemaakt en beheerd door bedrijven die veel dienstverlening bieden.
- Scenario 2: De brave nieuwe wereld. Datasystemen centraal bij de staat om de maatschappij te beschermen en besturen.
- Scenario 3: Een prettig profiel. Autonomie bij burgers, doel van databeheer is winstoogmerk.
- Scenario 4: Ubuntu. Autonomie bij vele spelers en het doel is de gemeenschap te laten bloeien.



Scenario 1: Centraal Kapitaal

In het eerste scenario neemt Mascha ons mee naar een toekomst waarin mensen hun digitale bestaan massaal delen, terwijl ze zelf weinig of geen invloed hebben op de netwerken of platforms waarop dit gebeurt. De infrastructuur kan wel degelijk decentraal zijn, maar het beheer ervan ligt in handen van een centrale organisatie die winstoogmerk als doelstelling heeft. Denk hierbij aan een organisatie als LinkedIn of Google, die iedere gebruiker gepersonaliseerde diensten aanbiedt, maar wel centraal georganiseerd is. De organisatie staat op decentraal niveau interactie toe en moedigt gebruikers aan om hun profiel uit te breiden en aan te vullen, maar belooft de gebruiker niet financieel. Mensen bezitten geen data, geen technologie, en bezitten in dit scenario zelfs geen huizen, meubels, auto's, of kleding. In deze service-georiënteerde economie ligt de verantwoordelijkheid voor eigendom bij bedrijven, die klanten proberen te overtuigen om te investeren in hun waren en niet in die van de concurrent.

Gebaseerd op je dataprofiel – waarbij zaken als inkomen, status, waardering, relaties, geschiedenis, en talenten meewegen – worden diensten aangeboden en kunnen tegen betalingen in de vorm van geld of data bepaalde 'levensfasen' geïnitieerd worden. Slimme agenten zijn onderdeel van de dienstverlening en intelligentie kan gekocht worden door diegenen die dit kunnen betalen. Technologie is alom aanwezig, sensoren zijn zo klein en onopvallend dat je maar hoeft te zeggen wat je wilt hebben of het kan

gerealiseerd worden. Algoritmes die aan gezichtsherkenning, gedragsherkenning, stemmingsmeting, en gedachtevoorspellingen doen, analyseren je gedrag en kunnen dit met elkaar delen. Aan de hand van deze data kunnen zelfs digitale kopieën van personen gemaakt worden waardoor de grens tussen werkelijkheid en virtuele wereld steeds verder vervaagt. Een vraag die hierbij opkomt, is wat de rol van de echte wereld nog is als deze zo goed nagebootst kan worden.

Scenario 2: Centraal Collectief

In het datingscenario waar Paul op weg is naar een restaurant speelt de staat een sterke rol. Hij krijgt te maken met een predictive-analytics systeem dat niet alleen maar voorspelt, maar ook overgaat op automatische besluitvorming en restricties oplegt (zoals het verbieden dat Paul de auto bestuurt). Het is een wereld waarin de overheid veel vertrouwen heeft in haar eigen data en op basis hiervan zelfs handelt als een overtreding niet heeft plaatsgevonden. De ingrepen zijn over het algemeen mild wat leidt tot een beeld van een Big Nudging-society, waarbij burgers een 'autocorrectie' of corrigerend tikje ontvangen als ze buiten de lijntjes kleuren. Tegelijk wordt er proactief gestuurd op goed gedrag en ideale burgers; de maakbare samenleving wordt een nieuw ideaal dat kan worden waargemaakt door de inzet van datatechnologie.

Een ander ideaal van de staat is om te komen tot personalisering; men is niet zomaar een burger, iedereen kan persoonlijk beoordeeld, beloond, gekend en gestraft worden. Hiervoor is het nodig alles in de publieke ruimte te monitoren, maar ook uitgebreide persoonlijke dataprofielen van burgers op te stellen en mee te wegen. Meetapparatuur van de staat neemt verschillende vormen aan; een auto kan ook als medische check-up fungeren.

Een vraagstuk dat in dit scenario op de achtergrond speelt is of de idealen van de staat aansluiten bij de idealen van burgers. Hoe zou het systeem reageren op burgerlijke ongehoorzaamheid? Zou een Rosa Parks in een hoogtechnologische tijd kunnen demonstreren door een bus in te stappen waar ze volgens de regels niet in mag? Of auto-corrigeert het systeem haar voordat het zover komt? Hebben burgers zelf de vrijheid en ruimte om de brave samenleving bij te sturen en te experimenteren met andere samenlevingsvormen?

Scenario 3: Decentraal Kapitaal

Sharon en Lani presenteren in scenario drie een toekomst met een gedistribueerd data-model. Een aantal van de Peer2Peer-initiatieven die we zien opkomen werken op een vergelijkbare wijze. Volgens Kostakis en Bauwens (2014) proberen ze meer mensen te betrekken bij handel en streven ze naar meer gelijkheid en participatie op de markt. Het kan gezien worden als een beweging die zich tegen de dominantie van grote spelers verzet, terwijl ze wel vasthoudt aan kapitalistische principes.

Computerfilosoof en klassiek musicus Jaron Lanier (waarnaar de namen Sharon en Lani impliciet verwijzen) schetst een soortgelijk toekomstscenario waarin we zelf bepalen wie onze data mag gebruiken en tegen welke kosten (Lanier 2014). Zo zouden we naar een economie kunnen streven die mensen betaalt voor de gegevens die ze online zetten. Er zijn steeds meer initiatieven om dit soort toekomst mogelijk te maken. Bitcoin is een voorbeeld van een systeem dat vandaag de dag al op zo'n decentrale manier werkt. Ook crowdfunding-platforms zoals Kickstarter of AirBnB zijn voorbeelden waarbij mensen decentraal geld kunnen genereren.

Op datavlak zijn er een aantal initiatieven zoals Datacoup of CitizenMe die beloven dat individuen hun eigen dataprofiel kunnen opbouwen, beheren en verkopen aan partijen die aan hen worden voorgesteld. "If your data is the new commodity in a digital world, CitizenMe welcomes you into the marketplace", schrijft het bedrijf op haar website. De Nederlandse student Shawn Buckles zette zijn persoonlijke data te koop via een veilingsite en ontving er 350 euro voor. Critici zoals Byun Chul Han zijn skeptisch of dit werkelijk bijdraagt aan een gezonde samenleving: "Zo wordt alles totaal geëconomiseerd. De mens heeft een akker nodig waar niet direct geogost moet worden, een speelruimte waar je je ook kan ophouden zonder winst te moeten generen" (In Van Verschuer 2014). De vraag van wie welke data zijn is een vraag die de komende jaren beantwoord moet worden.

Scenario 4: Decentraal Collectief

Jins scenario neemt concreet afscheid van een marktsysteem zoals we dit kennen en is omschreven als een transitiefase naar een open source-samenleving waarin digitale diensten zonder verdienmodel bestaan en gratis zijn voor leden van de gemeenschap. Men bouwt aan een nieuwe wereld terwijl de oude nog bestaat. Mensen van buiten de Ubuntu-gemeenschap kunnen gebruikmaken van Ubuntu-diensten, maar betalen hier wel voor. Het idee van Ubuntu is dat de gemeenschap centraal staat en dat alle data die geproduceerd worden binnen de gemeenschap van iedereen is. Dit geldt niet alleen voor de data, maar ook voor de architectuur, de software, de hardware en alles wat binnen het Ubuntu-domein gemaakt wordt. Men opereert wel autonoom en maakt keuzes over eigen data en bestaan, maar de gemeenschap speelt bij deze keuzes een belangrijke rol. Jin hint erop dat innovatie, vernieuwing en standaardisatie minder ontwikkeld zijn dan in een wereld waarin data centraal worden verzameld en ingezet. Vanwege het decentrale karakter zijn er vele initiatieven die aan elkaar gekoppeld zijn en zo een mondiaal ecosysteem vormen.

Conclusies: tijd voor decentralisatie?

Het vier-scenariomodel is een veelgebruikte methode om toekomst te verkennen en voor beleidsmakers en strategen vraagstukken bloot te leggen en discussies over de toekomst te faciliteren. Gebruikmakend van dit model gaat dit hoofdstuk in op vier toekomstscenario's waarbij de vragen hoe data georganiseerd worden (centraal of decentraal) en met welke reden data geproduceerd worden (met winstoogmerk of gemeenschapsgericht) leidend zijn. De vier scenario's presenteren alternatieve datamodellen waarin kwesties over data-eigenaarschap, systeemontwerp, databeheer en bestuur fundamenteel verschillen. De scenario's tonen aan dat data, afhankelijk van de ontwerpkeuzes die we maken, op heel verschillende manieren waarde kunnen creëren. Het kiezen van alternatieve datamodellen zal andere effecten hebben op de manier waarop we ons tot elkaar verhouden en waarop we de openbare ruimte, onze samenleving en ons bestuur inrichten.

Op dit moment kunnen de meeste Big Data-toepassingen in het kwadrant 'centraal kapitaal' geplaatst worden, wat inhoudt dat data worden geproduceerd met winstoogmerk door bedrijven die deze centraal organiseren. Vanwege Big Data's honger naar meer data en de wens data uit verschillende datasets te combineren zien we steeds meer adoptie van open source infrastructuur en open data binnen dit model, maar het winstoogmerk blijft leidend en het is opvallende dat modellen waarin decentrale organisatieprincipes leidend zijn nog in de kinderschoenen staan. Voordelen van dit soort modellen zijn dat ze een hoge mate van betrokkenheid, diversiteit en inclusiviteit kennen, transparantie hoog in het vaandel hebben staan, minder leiden onder het schrikbeeld van een 'Big Brother' die mensen stiekem in de gaten houdt, en individuen intrinsiek motiveren door ze beslissingsbevoegdheid en autonomie te geven. Het is goed mogelijk dat we een toekomst tegemoet gaan waarin we scenario's zien die bruggen slaan tussen de vier kwadranten, maar de strikte indeling van de vier kwadranten helpt scherp te krijgen welke spanningsvelden er zijn tussen verschillende datamodellen.

Data en de maakbare samenleving: realiteit of fabel?

In *smart cities* kan in de toekomst iedere beweging, ieder voorval en elke activiteit waargenomen en vastgelegd worden. De opkomst van *datafication*, het meten van alles wat los en vast zit, maakt het steeds meer mogelijk om data in te zetten voor het ontwerpen van onze ideale samenleving. Dit biedt ongekende mogelijkheden voor ingenieurs bij het ontwerpen van artificiële producten en processen. Herbert Simon, Amerikaans psycholoog en socioloog die wel wordt gezien als de vader van de moderne bedrijfskunde, ziet het ontwerpproces als een probleemoplossend proces, gericht op het oplossen van maatschappelijke problemen. Hedendaagse en toekomstige technologie brengt mensen en machines steeds dichter bij elkaar.

Een datagedreven samenleving leidt tot datagedreven besluitvorming. De samenleving wordt daarmee rationeler en maakbaarder. Bestuurders van deze wereld worden bijgestaan door systemen die draaien op kunstmatige intelligentie die geprogrammeerd is om de wereld te verbeteren, en die veel besluiten kunnen overnemen. De kunstmatige intelligentie regeert als een president met wijsheid voortkomend uit data: nooit meer oorlog, honger of vervuiling en op weg naar wereldvrede. In zo'n technocratische kijk op de wereld wordt ons dagelijks bestuur vervangen

door algoritmes die op basis van data beslissingen nemen.

In onze dagelijkse wereld zijn de eerste tekenen van *technocratic governance* al zichtbaar, met name op het gebied van smart cities, fraudedetectie en veiligheid. Bij een aanvraag voor een visum voor de Verenigde Staten krijg je binnen zeer korte tijd een antwoord. Algoritmes bepalen of je aanvraag akkoord is en ook of je potentieel gevaarlijk bent. Dit soort ontwikkelingen leidt tot het geloof in de maakbaarheid van de samenleving.

Algoritmes die bepalen wat goed en fout is, zijn niet per se op logica gebaseerd. Uitkomsten zijn niet altijd herleidbaar, omdat ze niet gebaseerd hoeven te zijn op oorzaak-gevolg-verbanden. Google baseerde voorspellingen van griep-epidemieën op zoekopdrachten, wat goed werkte totdat de griep een keer vroeger dan in andere jaren kwam. Dit leidde tot veel media-aandacht, waardoor mensen hierop gingen googelen, en de algoritmen op basis daarvan (opnieuw) met verkeerde voorspellingen kwamen (Butler 2013). Logica en algoritmes zijn verschillende dingen. Data en bijbehorende algoritmes geven een bepaalde kijk op de werkelijkheid, die niet overheen hoeft te komen met de werkelijkheid. Constant evalueren en bijstellen door mensen is dus nodig.

De rol van mensen is daarom belangrijk bij het gebruik van data. Ook de omgeving, de voorkeuren en de beperkte cognitieve mogelijkheden van de besluitvormer spelen een rol bij het maken van keuzes. Mensen valt het op als bepaalde patronen niet passen in een situatie. Maar mensen hebben cognitieve beperkingen, wat Simon *bounded rationality* noemt (1978). En als mensen kunnen waarnemen wanneer patronen niet logisch zijn, waarom ontwerpen we dan geen algoritmes die dit ook kunnen – en dan nog beter ook?

Voor het eindresultaat zijn zowel de data als de algoritmes waarmee de data bewerkt en gebruikt worden

van belang. Veelal worden data als feitelijk gezien en algoritmes als gezaghebbend. Maar klopt dit wel? Data kunnen gemakkelijk gemanipuleerd worden, sensoren kunnen slecht werken, tijdens het versturen kan iets misgaan met de data, ze kunnen onderschept of verdraaid worden. Grote gevaren liggen op de loer. Algoritmes vereisen bepaalde input en kunnen vaak niet omgaan met uitzonderingen en nieuwe ontwikkelingen omdat deze nog niet 'in het systeem zitten'. Dit soort zaken stelt grenzen aan de maakbaarheid van de samenleving.

Dat is een van de redenen waarom het openen van data en de bijbehorende algoritmes nodig is om te begrijpen hoe de uitkomsten geïnterpreteerd moeten worden: meer transparantie, meer verantwoording.

Deze kunnen dan gecontroleerd worden, fouten kunnen gedetecteerd worden en misbruik kan zo voorkomen worden. Meer transparantie en verantwoording.

Maar data is moeilijk te overzien, algoritmes zijn nauwelijks te begrijpen en het gebruik ervan leidt juist tot minder transparantie en verantwoording. Waar kan de oplossing vandaan komen? We zoeken die in automatisering waarbij algoritmes andere algoritmes controleren. Het vroeg meenemen van de eisen uit de samenleving in het ontwerpproces, het maken van afwegingen en oog

De kunstmatige intelligentie regeert als een president met wijsheid voortkomend uit data: nooit meer oorlog, honger of vervuiling en op weg naar wereldvrede.

hebben voor wat de samenleving echt wil is belangrijk. Participatie, het transparant maken van afwegingen, en het evalueren van consequenties zijn belangrijke stappen in het ontwerpen van de datagedreven samenleving van morgen.

Welke rol speelt Big Data in de maakbaarheid van de samenleving van de toekomst? Ontwikkelingen lijken tot een grotere maakbaarheid te leiden en bestuurders streven naar steeds meer controle. De maakbaarheid van de samenleving lijkt groter te worden, maar daar zijn beperkingen aan zolang mensen een grote rol blijven spelen. Wat zijn de grenzen van de beperkingen? En hoe ver zijn de beperkingen op te rekken om tot een big brother-achtige samenleving te komen met kunstmatige intelligentie als president?

Persoonsgegevens als kans en bedreiging voor Big Data

Big Data biedt enorme economische en maatschappelijke kansen. Alleen al het in goede banen leiden van grotere verkeersstromen van al dan niet zelfrijdende voertuigen en het oplossen van ziekten op basis van nu nog onwaarschijnlijke correlaties verbetert ons leven. Maar juist bij die menselijke aspecten dreigt Big Data een synoniem te worden voor privacyschending. Of dat terecht is, is twijfelachtig. Big Data in combinatie met het gebruik van kunstmatige intelligentie kan grote voordelen bieden en soms zelfs de privacy verder helpen.

Slimmer omgaan met gegevens

Neem het digitaliseren van medische dossiers. Op dit moment is er een enorme druk om over te gaan op een tussen zorgverleners uitwisselbaar Elektronisch Patiëntendossier (EPD). De aanpak om de ouderwetse kaartenbak van de arts te digitaliseren is niet vernieuwend. Om privacyrisico's te beperken, wordt arbitrair een grens getrokken op een maximaal toegankelijke historie van vijf jaar. Daardoor blijft soms relevante informatie buiten beeld. De kwaliteit van de dossiers wordt verder aangetast door te schiften tussen soorten informatie die wel en niet uitwisselbaar zijn. De gedachte is en blijft dat de medische data tussen de zorgverleners heen

en weer moeten. Dat maakt zo'n systeem kwetsbaar. In veel situaties is het uitwisselen van een heel dossier echter niet noodzakelijk. Big Data kan dan een uitkomst bieden door veel beperktere informatie te geven in antwoord op concrete vragen: mag een bepaald medicijn worden verstrekt en is dat het meest optimale voor deze patiënt? Voor welke patiënten is een nieuw medicijn nuttig? Voorwaarde is dat informatie uiteindelijk ergens centraal bij elkaar komt.

Een iets andere kijk op medische gegevens geeft dossiers juist meerwaarde. Dankzij Big Data is het mogelijk om niet benodigde informatie weg te houden bij derden, terwijl de kwaliteit van zorg verbetert. Een systeem gebaseerd op 'nice to know' kan veranderen in een systeem op basis van 'need to know'. Ook privacyzorgen rond zorgverzekeraars kunnen vermeden worden. Nu besluit een zorgverzekeraar op basis van GGZ-declaraties regelmatig tot aanvullend onderzoek. Dan komen gevoelige dossiers op tafel en komen er discussies over nut en noodzaak van een behandeling. Dit om fraude tegen te gaan. Big Data kan op basis van diagnoses en totaalbeelden deze nut-en-noodzaakdiscussie signaleren zonder verzekeraars inzage in gevoelige dossiers te geven. Alle betrokkenen (patiënt, zorgverlener en verzekeraar)

hebben vooraf een beeld van het behandelplan en de kosten. Dat vereist een beleidsmatige aanpassing om grenzen aan dekking vooraf goed te benoemen. Maar het voordeel is dat persoonsgegevens vertrouwelijker worden behandeld.

Zo zijn er veel voorbeelden te verzinnen. Een van de grootste inbreuken op de persoonlijke levenssfeer op dit moment is de inzet van beveiligingscamera's. Vooral de politie werkt er hard aan om toegang te krijgen tot zoveel mogelijk beeldmateriaal om hun werk te ondersteunen. Bij het kijken naar de beelden komen veel mensen in beeld, die daarna onnodig in de gaten worden gehouden. Inmiddels zijn er al diverse bedrijven die op basis van het analyseren van grote hoeveelheden camerabeelden afwijkend gedrag detecteren. Of die het mogelijk maken om te zoeken op bepaalde

kenmerken. Een voorbeeld van zo'n zoekactie betreft personen of objecten die tussen bepaalde tijden van punt A naar punt B

zijn gegaan. Uit de hoeveelheid gegevens wordt een samenvatting gemaakt van alles wat aan de criteria voldoet. Duizenden uren beeld kunnen op die manier worden samengevat in soms enkele seconden, waarbij degene die kijkt geen andere personen te zien krijgt. Het relevante stuk beeld kan geanalyseerd worden, waardoor zeer minimalistisch in de gegevens wordt gezocht. Het selectiever zoeken levert winst voor de persoonlijke levenssfeer en ook voor de tijdbesteding van de politie. Win-win. Maar het blijft een risico dat mensen met dit soort technieken zonder aanleiding intensiever worden gevolgd. Goede, heldere en strikte regelgeving kan dit verhinderen.

De mens als eigenaar van zijn gegevens

Persoonsgegevens zijn wettelijk heel breed gedefinieerd als alle gegevens die betrekking hebben op, of herleidbaar zijn tot een natuurlijk (levend) persoon. Om zulke data te verwerken moet er sprake zijn van toestemming, een vitaal belang of een wettelijke basis. Vanaf mei 2018 worden regels aangescherpt. Alles lijkt goed geregeld te zijn op het Europese continent. Maar juist Big Data is de tool om deze regelgeving eenvoudig te omzeilen. Door een database zo aan te passen dat gegevens niet meer te herleiden zijn tot een persoon is er namelijk geen sprake meer van persoonsgegevens en mag je de datasets zonder problemen combineren en exporteren. Het kopiëren van gegevens is in het internettijdperk snel gedaan en nadat de data met allerhande bronnen verrijkt zijn worden het weer per-

Het punt is niet of je iets te verbergen hebt, maar of je vrij kunt beschikken over datgene waar je recht op hebt.

soonsgegevens, maar nu in een andere juridische omgeving waar deze data *wel* te gebruiken zijn. De betrokkene (zoals de juridische aanduiding is van de persoon om wie het gaat) heeft hierbij het nakijken. We zullen dus anders naar data moeten kijken. In mijn boek *Digitale Stormvloed* kom ik tot de conclusie de oude manier van kijken naar persoonlijke data daarom niet werkt. Het zou logisch zijn om de eigendomsvraag centraal te stellen en mensen als rechthebbende op de eigen gegevens te beschouwen. Want behalve dat persoonsgegevens onze belangen aangaan, vertegenwoordigen ze ook een waarde.

In veel gevallen verstrekt de betrokken

persoon de data zelf door informatie te geven of door gedrag te laten registreren. Het gebruik van die data door bedrijven, overheden en anderen zou middels een licentie horen te gebeuren, waarbij het bedrijf het nodige mag, maar waar ook beperkingen voor gelden. En hoe die licentie wordt verleend hoort de individuele keus te zijn van de betrokken persoon. Want wat er met gegevens gebeurt, wat de mogelijke gevolgen zijn, en vooral hoe die consequenties de betrokkene kunnen raken, is immers relevant voor de persoon zelf. In de huidige structuur van onze informatiesamenleving is dat nog niet geregeld. We geven nu (vaak impliciet) toestemming voor gebruik zonder te weten waarvoor de gegevens precies gebruikt worden.

Wanneer data rechten zouden geven aan degene die het betreft, verandert de discussie radicaal. Natuurlijk zouden we er altijd voor kunnen kiezen onze gegevens in licentie te geven aan een bedrijf dat ze voor ons te gelde kan maken, maar dan is het gebruik overzichtelijk, het model zakelijk en is er sprake van eerlijke handel. Het is dan duidelijk wat het gebruiken van de dienst echt kost.

Het kwantificeerbaar maken van persoonsgegevens ontkracht statements als 'ik heb niets te verbergen' of 'je hebt wel iets te verbergen'. Het punt is niet of je iets te verbergen hebt, maar of je vrij kunt beschikken over datgene waar je recht op hebt. Iemand die stelt 'niets te verbergen te hebben' wekt de suggestie dat zijn data onbeperkt gebruikt mogen worden en dat

het uitbaten van zijn gegevens hem of haar niet raakt. Het valt te betwijfelen of die keuze weloverwogen is, maar centraal staat dat iedereen zelf bepaalt wat er met zijn eigendom gebeurt. Dat iemand geen waarde aan zijn gegevens hecht, verandert niets aan de eigendomsvraag. Eigendom is immers absoluut en kan niet zomaar vervreemd worden. Als de overheid voor een breder maatschappelijk belang persoonlijke gegevens gebruikt, hoort daar een kloppende procedure bij. Zo'n inperking zorgt er ook voor dat de gegevens niet voor vage, onduidelijke doelen worden gebruikt. Een 'betere dienstverlening', 'de veiligheid' of 'maatschappelijke belangen' voldoen dan niet.

Alles lijkt goed geregeld te zijn op het Europese continent. Maar juist Big Data is de tool om deze regelgeving eenvoudig te omzeilen.

Technisch gezien is het regelen van toestemming en het maken van afspraken niet zo ingewikkeld als het lijkt. Het onderzoek *Licht op de Digitale Schaduw*, geschreven door diverse experts in opdracht van het Ministerie van Economische Zaken, laat duidelijk zien dat mensen prima bereid zijn hun persoonsgegevens af te staan voor Big Data-verwerking als ze te horen krijgen wat ermee gebeurt. Het vertrouwen neemt zelfs toe en opent ook zakelijke mogelijkheden. Helaas zal het van het bedrijfsleven afhangen of ze kunnen zien in het gebruik van Big Data met respect voor de rechten van mensen. Daar ligt een gouden toekomst. Niet in de datahonger naar ieders privéleven.



1995. 11

1994. 11

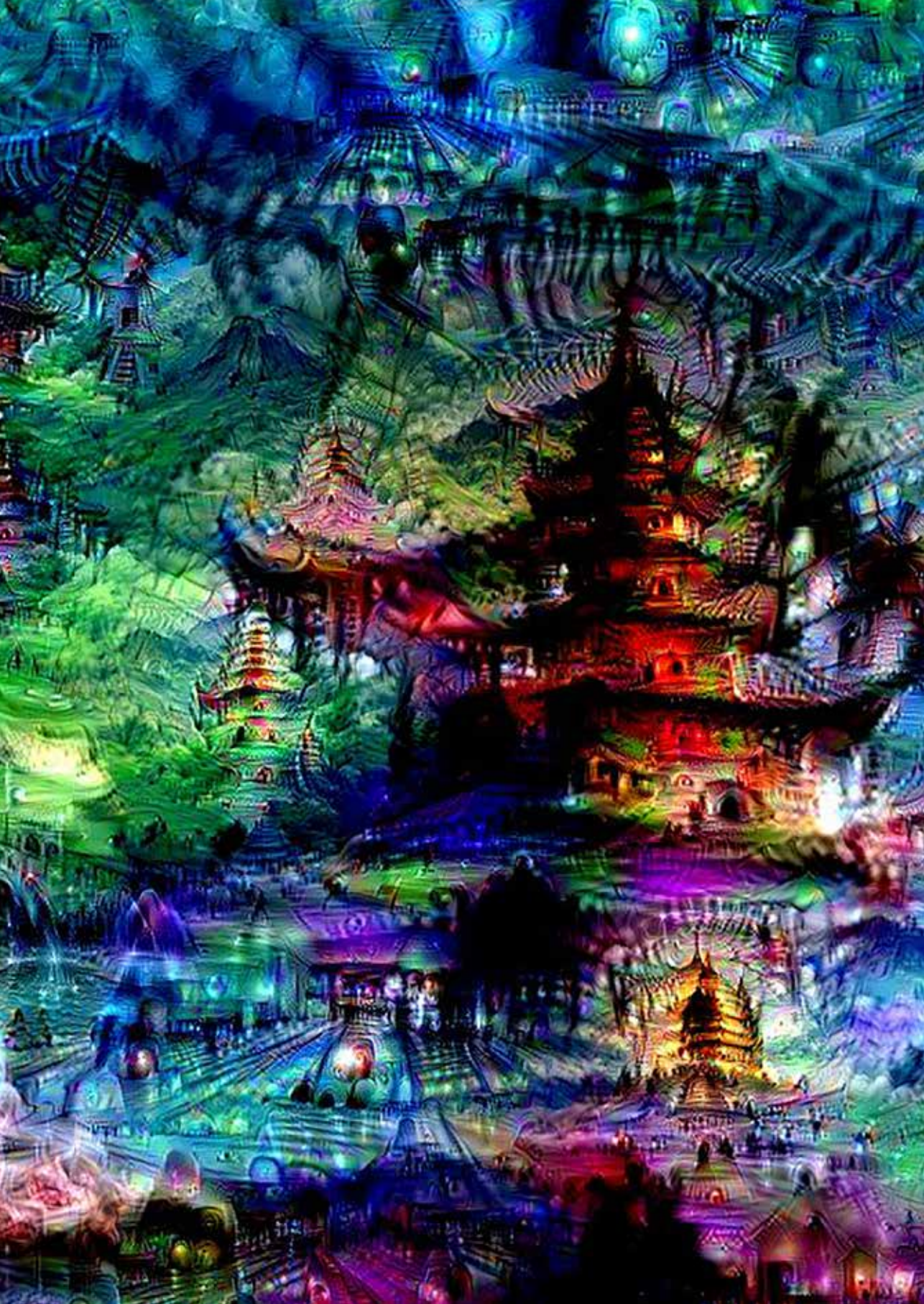
1992 (11)

(11)

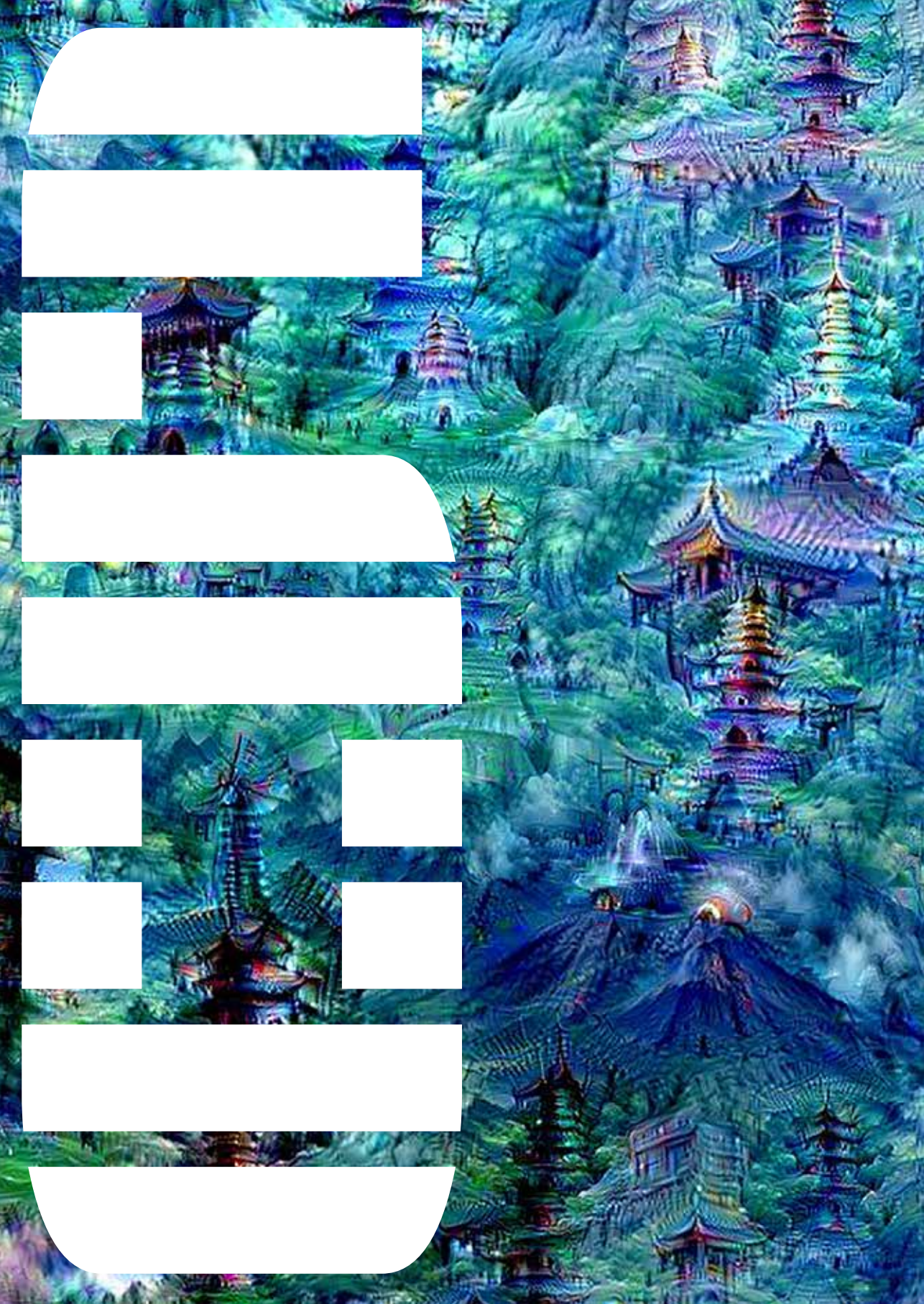
2002 (11)

2002 (11)

2003 (11)







6. WAT NU?

Het laatste hoofdstuk van een verkenning biedt vaak een vooruitblik en beweegt van 'hoe het is' naar 'hoe het zou kunnen zijn'. In deze publicatie doen we dit andersom en bewegen we juist vanuit 'hoe het zou kunnen zijn' terug naar 'hoe het nu kan'.

Filosoof Hans Achterhuis merkt in zijn boek *Koning van Utopia* op dat een utopie iets totalitairs kan hebben (2016). Utopieën worden opgesteld als reactie op imperfecte samenlevingen, maar de blauwdruk van een nieuwe samenleving mondt vaak uit in een soort 'controlestaat' waar alles 'volgens de letter moet'. Dit gebeurde bijvoorbeeld met de utopieën van Thomas More en Karl Marx, op diverse momenten en plaatsen.

Een les die we hieruit kunnen trekken is dat we niet op zoek moeten gaan naar een nieuw paradigma. De kritiek is een oproep tot het 'kleine verhaal', de heterotopia zoals Foucault het noemde, of een kader voor meerstemmigheid. Het is een oproep om een samenleving te realiseren waarin verschillende utopieën naast elkaar kunnen bestaan en mensen verschillende toekomstbeelden kunnen nastreven. In zekere zin is dit ook de opzet die we gekozen hebben bij deze publicatie. We hebben, met verschillende stemmen uit verschillende sectoren niet één visie op de toekomst proberen te realiseren, maar hebben talloze toekomstbeelden naast elkaar gezet. In deze meerstemmigheid zijn wel degelijk rode draden te vinden en die staan in dit afsluitende hoofdstuk centraal.

Kunstmatige intelligentie voor de meerderheid

De eerste rode draad stelt dat in het Big Data-veld bijzondere behoefte is aan meerstemmigheid. Niet alleen in de discussie over het fenomeen, maar ook in de ontwerpfase, op bestuursniveau, in het aanstellen van ethische commissies en op het niveau van de eindgebruiker. Zoals ook bij andere technologie het geval is zullen datasystemen van de toekomst de waarden van hun ontwerpers reflecteren. Marijn Janssen stelde in zijn bijdrage dat het vroeg meenemen van eisen uit de samenleving hierbij essentieel is: "participatie, het transparant maken van afwegingen, en het evalueren van consequenties zijn belangrijke stappen in het ontwerpen van de datagedreven samenleving van morgen." Als we dit niet doen riskeren we dat machines, misschien wel heel weldenkende machines, een afspiegeling zijn van een klein gedeelte van de samenleving en bij anderen juist weerstand oproepen. Zoals KI-professor Emma Brunskill zegt: "AI should benefit the many, not the few." Het doel van onderzoek naar KI zou dan ook niet moeten zijn om zo snel mogelijk KI te ontwikkelen, maar om KI zo te ontwikkelen dat het de samenleving ten dienst staat.

Dirk Helbing stelt dat diversiteit ook van belang is omdat we met de digitale wereld van morgen met onvoorstelbare complexiteit te maken zullen hebben. Net als bij een kathedraal met complexe ornamenten, is er niet een enkel perspectief dat het mogelijk maakt de schoonheid in zijn totaal te zien. Geen mens is slim genoeg om de complexiteit van de hyper-verbonden wereld van morgen te begrijpen en dat geldt volgens Helbing ook voor KI-systemen. Het idee van *singularity* heeft het gevaar in zich dat het tot een mislukte utopie leidt. We zouden juist het pluralisme, wat we in onze samenleving als belangrijke waarde bestempelen, in onze systemen moeten inbouwen. We zijn gebaat bij de wisdom of crowds, ook als de crowd voor een gedeelte uit KI-systemen bestaat.

Mens centraal

De mens, burger, of gebruiker, heeft een intrinsiek zwakke positie in de toekomstige informatiesamenleving. Datatechnologie brengt een bepaalde onbalans met zich mee omdat een individu niet voldoende kapitaal heeft om een datasysteem te ontwikkelen of op te zetten. Het is van belang om deze onbalans recht te trekken en de informatiepositie van mensen te bekrachtigen. Geef mensen controle over hun gegevens, over systemen, laat ze mee-ontwerpen, meetoetsen en betrek ze in discussies. Het is gemakkelijker om ongezien data over personen te verzamelen, ze door middel van cookies te volgen, en ze te verleiden om hun data af te staan, dan om de directe discussie aan te gaan, maar betekenisvolle instemming blijft de essentie van respect voor het individu en diens privacy. Als we privacy, transparantie, autonomie en identiteitsbescherming niet meenemen in de ontwikkeling van de datasystemen van morgen, resulteert dat in een ongezonde machtsbalans. Bovendien laten de scenario's zien dat naast privacyschendingen Big Data ook andere mogelijke schendingen – zoals risico op discriminatie, aantasten van de vrijheid van meningsuiting, en vrijheid van geloof en levensovertuiging – kunnen plaatsvinden. Hier zou in het datadebat van vandaag en morgen meer aandacht voor mogen zijn.

“Als je niet betaalt voor een dienst ben jij niet de consument maar het product”, stelde Andrew Lewis. Een kritiek punt in onze relatie met organisaties die data over ons vastleggen, is dat men niet goed begrijpt dat onze data verhandeld worden, hoe dit gebeurt en wat de waarde van onze data is. De in deze publicatie gepresenteerde machts-scenario's presenteren alternatieve datamodellen die laten zien dat decentralisatie van datadiensten zowel materiele als immateriele waarde zouden kunnen opleveren voor eindgebruikers en de samenleving.

Kennisbescheidenheid

Met de opkomst van datatechnologie die rekenkracht, opslag en algoritmische nauwkeurigheid maximaliseert krijgen we nieuwe gereedschappen om naar de wereld te kijken en daarmee een goudmijn aan kansen om tot meer kennis te komen. Toch is er

een noodzaak voor kennisbescheidenheid. Zelfs als we met duizenden data-elementen – waaronder genetische informatie, locatiedata en een goed gevuld profiel – conclusies willen trekken over mensen moet de mogelijkheid openblijven dat de analyse niet juist is. “Een fatsoenlijke samenleving laat mensen voor een belangrijk deel het verhaal over zichzelf vertellen”, schrijft Jeroen van den Hoven. Rogier Creemers vult aan dat we moeten beseffen dat we niet alles in cijfers kunnen uitdrukken. Er zijn zaken die in principe onmeetbaar zijn.

Bewust van onbewuste intelligentie

Hoewel de realisatie van singularity of van superintelligentie volgens de meeste KI-experts verre toekomstmuziek is, zien we de komende jaren overal om ons heen nieuwe KI opduiken. Sprekende huiskamers, denkende straatverlichting en persoonlijke digitale assistenten staan in de virtuele rij om u van dienst te zijn. Zelflerende systemen zijn op komst en het komt nu al voor dat zelfs de bedenkers van systemen niet exact begrijpen hoe bepaalde output tot stand is gekomen. We hebben te maken met steeds autonomer wordende intelligente machines. We hebben inzicht nodig in statistiek, waartoe Mireille Hildebrandt oproept in haar bijdrage, en moeten ook inzien dat we te maken krijgen met nieuwe typen organismes: non-organische intelligente wezens. Misschien hebben deze type wezens een nauwe vorm van intelligentie, maar ze zijn wel erg bekwaam in bepaalde handelingen. Denk aan de bijdrage van Emma AI, of het winnen van Go-bang door Alpha-Go. Naar verwachting krijgen we steeds meer te maken met dit soort intelligente modules, en het is belangrijk hierop te anticiperen, bijvoorbeeld door in te zien dat er een verschil is tussen wiskundige verbanden en het soort verbanden dat wij zelf voortdurend leggen. “Je moet een gezonde weerstand ontwikkelen tegen al die op data gebaseerde voorspellingen en een steekhoudend weerwoord kunnen geven als je getarget wordt door je robo (of door die van een ander)”, schrijft Hildebrandt.

Nederland, dataland?

Datagedreven technologie kan een belangrijke motor zijn voor economische groei en sociale vooruitgang. Hiervoor zullen het bedrijfsleven, het maatschappelijk middenveld, de overheid en de wetenschap moeten samenwerken om zowel aandacht voor de technologie als de daaraan gepaarde risico's te hebben. Volgens verschillende auteurs uit deze publicatie is er winst te halen voor Nederland door in te zetten op een koplopersrol in het datadomein. De spanning die nu ervaren wordt tussen Big Data à la Silicon Valley en onze eigen grondrechten zou een uitgangspunt kunnen bieden om deze rol te verzilveren. Jeroen van den Hoven doet een oproep tot maatschappelijk verantwoord innoveren met Big Data en stelt we in Nederland zouden moeten inzetten op *privacy by design*, en *privacy respecting of privacy enhancing technologies*.

Belang van de toekomst

Tot slot is het belangrijk dat we vooruit blijven kijken. De meeste plannen worden gemaakt op basis van inadequate toekomstbeelden. De problemen van het heden en het directe verleden worden hierbij geanalyseerd en geprojecteerd op de toekomst. Niemand kan voorspellen wat de toekomst zal zijn, maar er is een grote kans dat deze significant zal afwijken van het heden. Veel beleid, strategieën en plannen worden op dit moment gemaakt op basis van onwaarschijnlijke toekomst. Toekomstverkenningen kunnen van waarde zijn door uiteenlopende toekomst te presenteren en deze alvast te repeteren zodat we met nieuwe ogen naar de werkelijkheid leren kijken.

BIJLAGEN

Aanpak

Deze publicatie is gebaseerd op een toekomstverkenning die in de periode 2015-2017 door STT is uitgevoerd. De toekomstbeelden zijn tot stand gekomen door inzichten samen te brengen van meer dan 150 experts, belanghebbenden en geïnteresseerden die ten tijde van de verkenning actief waren op het gebied van data en de samenleving. Tijdens een reeks bijeenkomsten en met name de zogenaamde 'Datamore' workshops bij STT is aan de hand van foresight- en backcastingmethodes een serious game ontwikkeld om samen met deelnemers maatschappelijke uitdagingen, technologische trends en scenario's met betrekking tot Big Data en de toekomst in kaart te brengen. Aan de hand van interviews, focusgroepen en een online community is verdere input verzameld en output gedeeld om een voorstelling te kunnen maken van een leven met meer data.

Naast deze publicatie zijn diverse populaire en wetenschappelijke artikelen gepubliceerd, zoals:



Snijders, D., 2016. 'Wilde data: over de sociale gevolgen van Big, Open, en Linked Data systemen'. *Bestuurskunde*, 25 (1), 53-57.

Dwivedi, Y.K., Janssen, M., Snijders, D. *et al.*, 2016. 'Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modeling'. *Information Systems Frontiers*, 1-16.

Dwivedi, Y.K., Janssen, M., Snijders, D. *et al.*, 2015. Driving Innovation Using Big Open Linked Data (BOLD). *Open and Big Data Management and Innovation*. Springer, 3-9.

Nenciu, G., Duin, P., Janssen, M., Hulstijn, J., Snijders, D., 2016. *A BOLD Approach to Smart Cities*. Imagining the Future Conference, Istanbul.

Dankwoord en betrokkenen

De resultaten van deze verkenning waren niet mogelijk geweest zonder de deelname van een groot aantal personen en organisaties. Zonder hun inzet en wetenschappelijke, beschouwende of creatieve inbreng had u deze letters niet gelezen. Bij deze willen we iedereen die heeft deelgenomen aan de verkenning bedanken voor hun waardevolle inzichten.

In het bijzonder danken we de mede-auteurs die de uitdaging aandurfden om de verre horizon op te zoeken; de respondenten die tijdens workshops en interviews hielpen toekomstbeelden te schetsen; de stuurgroepleden die zich gedurende het traject zowel als klankbordgroep, ambassadeurs en denktank inzetten; en natuurlijk de collega's bij STT die altijd klaarstonden om de toekomst in het heden te vieren.

Samenstelling Stuurgroep

Jos Keurentjes, *voorzitter*, CSO TNO

Arre Zuurmond, Ombudsman Gemeente Amsterdam en omstreken

Eric van Tol, Directeur Expertisecentrum Big Data Fontys

Fred Herrebout, Strategy Manager T-mobile

Geert Jan Houben, Hoogleraar Big Data, KIVI / TU Delft

Gerard Smit, CTO IBM

Hans Timmerman, CTO Dell EMC

Jan Roggeveen, Business Development Cisco

Marieke van Lier Lels, Commissaris Eneco, NS, RELX en TKH

Marijn Janssen, Hoogleraar ICT & Governance, TU Delft

Max Remerie, Voormalig directeur Innovatie Siemens

Pierre Morin, Voormalig directeur STT

Steven Luitjens, Directeur Informatiesamenleving, Ministerie Binnenlandse Zaken

Willem Vermeend, Hoogleraar Economie 4.0, Open Universiteit

Samenstelling Projectgroep

Dhoya Snijders, projectleider

Annette Potting en **Japke Schreuders**, projectondersteuning

Georgiana Nenciu, masterstudent TU Delft en stagiair STT

Stichting Toekomstbeeld der Techniek

De Stichting Toekomstbeeld der Techniek (STT) is in 1968 opgericht door het Koninklijk Instituut Van Ingenieurs (KIVI). STT is een onafhankelijke stichting die gefinancierd wordt uit bijdragen van overheid en bedrijfsleven. STT voert brede toekomstverkenningen uit die domein-overstijgend en interdisciplinair zijn, op het snijvlak van technologie en samenleving.

Het Algemeen Bestuur (AB) van STT bestaat uit topmensen uit de overheid, het bedrijfsleven, de onderzoekswereld en uit maatschappelijke organisaties. Het AB denkt mee over de STT-programmering, is betrokken bij verkenningen en vormt een belangrijke denktank waarbinnen de bestuursleden praten over toekomstige technologische ontwikkelingen en innovatie.

De STT Academy ontplooit daarnaast diverse activiteiten, zoals het co-financieren van bijzondere leerstoelen, methodiekontwikkeling, de organisatie van masterclasses, en het beheer van het Netwerk Toekomstverkenningen en Young STT, met high potentials uit de deelnemende organisaties.

Informatie over STT, haar activiteiten en haar producten is te vinden op de website www.stt.nl

Publicaties

Een aantal recente verkenningen van STT:

STT85 Wie wij worden, Toekomstbeelden van mensen in 2050

Ellen Willemse, 2016

STT 84 Nationale Toekomstmonitor; Hoe kijken Nederlanders naar technologie en de toekomst

Dhoya Sniijders, 2016

STT 83 Een Oceaan vol mogelijkheden

Stephanie IJff, Marie-Pauline van Voorst tot Voorst, 2016

STT 82 Beter, Toekomstbeelden van technologie in de zorg

Ellen Willemse, 2015

STT 81 Van autonome robots tot zilte aardappels

Silke de Wilde, 2015



Algemeen Bestuur STT (januari 2017)

Dagelijks bestuur

- Ir. R. Willems**, *voorzitter*, Voormalig president-directeur Shell Nederland
Ir. C.C.J. Vincent MBA, *vicevoorzitter*, CEO Meratus Line
Mevr. dr. ir. N. Buitelaar MBA, *secretaris*, Chief Business Officer, BiosanaPharma
Ir. J.H.J. Mengelers, *penningmeester*, Voorzitter College van Bestuur TU Eindhoven
Drs. M.E. Remerie, *lid DB*, Voormalig directeur Business Development Siemens Nederland

Algemeen bestuur

- Prof. dr. E.H.L. Aarts, Rector Magnificus Tilburg University
H. Blokhuis, CTO & Director for Collaboration & Video, Cisco
Drs. A.C.J. van den Boom, Directeur Interpolis
Ir. T.J. Bosma, Program Director Power Systems & Electrification, DNV GL Strategic Research & Innovation
Mevr. L. de Bruijn, Directeur Nederland ICT
J.W. Dijkhoorn MSc, Director Analytical Platform CoE – SAS South West Europe
Drs. Th. N.M. Föllings, Manager Bedrijfsomgeving, Ontwikkeling & Innovatie Oost NV
Ir. B.C. Fortuyn, Lid Raad van Bestuur van Siemens Nederland NV
Drs. J.H. de Groene, Algemeen Directeur Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)
Dr. B. ter Haar (*waarnemer*), Directeur-generaal Participatie en Inkomenswaarborg, ministerie van SZW
Drs. F.P.U. Haffmans, Managing Director, Head of Corporate Banking Benelux, Bank of America Merrill Lynch
Dr. M. van den Hauten (*waarnemer*), Plv. directeur Onderzoek en Wetenschapsbeleid, ministerie van OCW
Drs. J.H. Heres (*waarnemer*), Hoofd Centrale Eenheid Strategie, ministerie V&J
F. Herrebout, Senior Strategy Manager T-Mobile
M.S. Hofland MSc, CEO Capgemini Consulting Nederland
Drs. A.J. van den Hoogen, Director R&D Products and Applications, Tata Steel Research, Development & Technology
Dr. H. van Houten, Executive Vice President, Philips Research
Dr. T. Jongsma, Directeur Stichting Public Private Partnership Institute for Sustainable Process Technology (ISPT)
Mevr. dr. M.J. Jonkman, Corporate director R&D Koninklijke FrieslandCampina
Prof. dr. ir. J.T.F. Keurentjes, CSO, lid Raad van Bestuur TNO
R.A. Kleiburg, COO ECN
Mevr. mr. B. van der Kolff MBA, Directeur, Lid Directieteam VvAA
Ir. G.A. Kroon, Algemeen directeur ARCADIS Nederland
Dr. B. Leeftink (*waarnemer*), Directeur-generaal Bedrijfsleven en Innovatie, ministerie van EZ
Mevr. drs. E.P.J. Lemkes-Straver, Algemeen Directeur ZLTO
Mevr. ir. M. van Lier Lels, Lid Raad van Commissarissen Eneco, NS, RELX en TKH; diverse bestuursfuncties
Prof. dr. F. Miedema, Decaan faculteit Geneeskunde Universiteit Utrecht, vicevoorzitter raad van bestuur UMC Utrecht
Ir. P.C. Molengraaf MBA, Voorzitter Raad van Bestuur Alliander
P.W. Mollema MSc, Director Environmental Management, Port of Rotterdam Authority, Ir. M.W. Moolhuijsen, Directeur Strategie Essent
Dr. ir. H.H. Nijhuis, Directeur STW
Mevr. prof. mr. A. Oskamp, Rector magnificus van de Open Universiteit
Prof. dr. ir. A. Osseyran, Managing director SURFsara
Mevr. ir. A.M. Ottolini, Algemeen directeur Evides Waterbedrijf
Ir. C.J. Rameau, Lid Raad van Bestuur Eneco Holding NV
Mevr. drs. M.V.I.M dos Ramos, Directeur KIVI
Ir. P. van Riel, CEO Fugro
Ir. P.W.F. Rutten MBA, Partner McKinsey and Company
Prof. dr. ir. W. van Saarloos, Vice-president KNAW
Mevr. drs. J.H. Scholten, Directeur VSNU
Ir. Y. Sebrechts, Executive Vice President Innovation, R&D, CTO Projects & Technology Royal Dutch Shell
Drs. R. Slotman, Directeur Regieorgaan SIA
Drs. ing. G.E.A. Smit, CTO IBM Benelux,

IBM Distinguished Engineer
F.E. Smith, Director Public Affairs, ANWB
Ir. M.T.J.H. Smits, Algemeen directeur Deltares
Prof. dr. ir. M. Steinbuch, Hoogleraar Automotive Technology TU Eindhoven
Mevr. dr. V.C.M. Timmerhuis, Algemeen Secretaris Sociaal-Economische Raad (SER)
Ir. H. Timmerman, CTO Dell EMC Nederland
Mevr. B. Tool, Directeur MaagLe-verDarmStichting (MLDS)

Dr. J.M.A. Verbakel, VP BCS & Foods Operations, Unilever
Prof. dr. M. Verkerk, Bestuurslid Vita-Valley, Bijzonder hoogleraar Christelijke Wijsbegeerte TU Eindhoven
Mevr. prof. dr. ir. M.P.C. Weijnen, Hoogleraar TU Delft, faculteit Techniek, Bestuur en Management, lid WRR
Mevr. mr. J.S. van der Woude, Company Secretary and Legal Director Continental Europe, RELX
Drs. R. Zandbergen, CEO USG People

Prof. dr. A.N. van der Zande, Directeur-Generaal Rijksinstituut voor Volksgezondheid en Milieu (RIVM)

PwC en DSM hebben momenteel geen afgevaardigden in het bestuur

Adviserende leden uit de STT-Academy

STT hoogleraren

Prof. dr. ir. V.A.W.J. Marchau, Radboud Universiteit: onzekerheid en adaptiviteit van maatschappelijke systemen

Mevr. prof. dr. M.H. Martens, Universiteit Twente: Intelligent Transport Systems (ITS) en Human Factors

Directeur bureau Dr. P.A. van der Duin

Literatuur

- Achterhuis, H., 2016. *Koning van Utopia*. Rotterdam: Lemniscaat Uitgeverij.
- Anderson, C., 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16 (7), 16–07.
- Angwin, J., Larson, J., Mattu, S., and Kirchner, L., 2016. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. *ProPublica*.
- Arbesman, S., 2011. Quantifying the ease of scientific discovery. *Scientometrics*, 86 (2), 245–250.
- Baller, S., Dutta, S., and Lanvin, B., 2016. The global information technology report 2016. In: *World Economic Forum, Geneva*. 1–307.
- Barr, A., 2015. Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms. *Wall Street Journal Blog*, 1 Jul.
- Bensinger, G., 2014. Amazon Wants to Ship Your Package Before You Buy It. *Wall Street Journal Blog*, 17 Jan.
- Berg, B. van den, 2010. *Ambient Intelligence: Wat, wie en... waarom?* Institute for Prospective Technological Studies (IPTS), 2, 3.
- Bezold, C., 2009. Jim Dator's alternative futures and the path to IAF's aspirational futures. *Journal of Futures Studies*, 14 (2), 123–134.
- Billao-Osorio, B., Dutta, S., and Larvin, B., 2014. The global information technology report 2014: Rewards and risks of big data. In: *World Economic Forum*.
- Bostrom, N., 2014. *Superintelligence: Paths, dangers, strategies*. OUP Oxford.
- Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A.M., Jozefowicz, R., and Bengio, S., 2016. Generating sentences from a continuous space. In: *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, 10–21.
- boyd, d. and Crawford, K., 2011. Six provocations for big data. In: *A decade in internet time: Symposium on the dynamics of the internet and society*. Oxford Internet Institute Oxford.
- boyd, d. and Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15 (5), 662–679.
- Broek, A. van den, Campen, C. van, Haan, J. de, Roeters, A., Turkenburg, M. and L. Vermeij, 2016. *De toekomst tegemoet*. Den Haag: SCP.
- Butler, D., 2013. When Google got flu wrong. *Nature*, 494 (7436), 155.
- Chalmers, D., 2010. The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17 (9–1), 7–65.
- Chui, M., Löffler, M., and Roberts, R., 2010. The internet of things. *McKinsey Quarterly*, 2 (2010), 1–9.
- Connolly, K., 2016. Angela Merkel: internet search engines are 'distorting perception'. *The Guardian*, 27 Oct.
- Courtland, R., 2015. Gordon Moore: The Man Whose Name Means Progress. *IEEE Spectrum: Technology, Engineering, and Science News*, 30 Mar.
- Cowls, J. and Schroeder, R., 2015. Causation, Correlation, and Big Data in Social Science Research. *Policy & Internet*, 7 (4), 447–472.
- Cox, M. and Ellsworth, D., 1997. Application-controlled demand paging for out-of-core visualization. In: *Proceedings of the 8th conference on Visualization'97*. IEEE Computer Society Press, 235–ff.

- Crawford, K. and Schultz, J., 2014. Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, 55, 93.
- Ducastel, K., 2001. ISTAG: Scenarios for Ambient Intelligence in 2010. IPTS (JRC).
- Duhigg, C., 2012. How Companies Learn Your Secrets. *The New York Times*, 16 Feb.
- Eggers, D., 2014. *The Circle*. 1st edition. New York: Vintage Books.
- Esguerra, R., 2009. Google CEO Eric Schmidt Dismisses the Importance of Privacy. *Electronic Frontier Foundation*, 10 Dec.
- Est, Q.C. van, Rerimassie, V., van Keulen, I., and Dorren, G., 2014. *Intieme technologie: de slag om ons lichaam en gedrag*. Den Haag: Rathenau Instituut.
- Evans, D., 2011. The internet of things: How the next evolution of the internet is changing everything. *CISCO white paper*, 1 (2011), 1–11.
- Expertgroep Big data en privacy, 2016. *Licht op de digitale schaduw. Verantwoord innoveren met big data*. Den Haag: Ministerie van Economische Zaken, rapport.
- Giaoutzi, M. and Sapio, B., 2013. In Search of Foresight Methodologies: Riddle or Necessity. In: M. Giaoutzi and B. Sapio, eds. *Recent Developments in Foresight Methodologies*. Springer US, 3–9.
- Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E.M., Sipos, B., and Birney, E., 2013. Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. *Nature*, 494 (7435), 77–80.
- Habegger, B., Hasan, O., Brunie, L., Bennani, N., Kosch, H., and Damiani, E., 2014. Personalization vs. privacy in big data analysis. *International Journal of Big Data*, 1 (1), 25–35.
- Han, B.-C., 2015. *The Transparency Society*. Stanford, California: Stanford Briefs.
- Hauswald, J., Laurenzano, M.A., Zhang, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R.G., Mudge, T., Petrucci, V., Tang, L., and Mars, J., 2015. Sirius: An Open End-to-End Voice and Vision Personal Assistant and Its Implications for Future Warehouse Scale Computers. In: *Proceedings of the Twentieth International Conference on Architectural Support for Programming Languages and Operating Systems*. New York, NY, USA: ACM, 223–238.
- Helbing, D., 2015. Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies. *Artificial Intelligence, and Manipulative Technologies (April 14, 2015)*.
- Helbing, D., Frey, B., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., van den Hoven, J., Zicari, R., and Zwitter, A., 2016. Digital-Manifest I u. II. *Spektrum der Wissenschaft*, 1, 51.
- Herbert, S.A., 1978. *The science of the artificial*. MIT Press.
- Hilbert, M., 2016. Big data for development: A review of promises and challenges. *Development Policy Review*, 34 (1), 135–174.
- Hirsch Ballin, E.M.H., Broeders, D., Schrijvers, E., van der Sloot, B., van Brakel, R., de Hoog, J., and others, 2016. *Big Data in een vrije en veilige samenleving*. Den Haag: WRR.
- Ivakhnenko, A., 1966. *Cybernetic predicting devices*. Washington: Joint Publications Research Service.
- Janssen, A., Kool, L., and Timmer, J., 2015. *Dicht op de huid: Gezichts-en emotieherkenning in Nederland*. Den Haag: Rathenau Instituut.
- Kalff, F.E., Rebergen, M.P., Fahrenfort, E., Girovsky, J., Toskovic, R., Lado, J.L., Fernández-Rossier, J., and Otte, A.F., 2016. A kilobyte rewritable atomic memory. *Nature Nanotechnology*, 11 (11), 926–929.


- Kennedy, J., 2016. *Big Data's Economic Impact* [online]. Committee for Economic Development of The Conference Board. [Accessed 26 Jan 2017].
- Klaassen, P., Schuijff, M., Smits, M., and others, 2009. *Mens van de toekomst-mens zonder toekomst: mensverbetering in cultureel, politiek en technologisch perspectief*. Den Haag: Rathenau Instituut.
- Kosinski, M., Stillwell, D., and Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110 (15), 5802–5805.
- Kostakis, V. and Bauwens, M., 2014. *Network society and future scenarios for a collaborative economy*. Springer.
- Kool, L., Timmer, J., and van Est, R., 2015. *De datagedreven samenleving.: Achtergrondstudie*. Den Haag: Rathenau Instituut.
- Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 70.
- Lanier, J., 2014. *Who owns the future?* Simon and Schuster.
- Lawler, R., 2011. *How Netflix Will Use Big Data to Push House of Cards* [online]. Available from: <https://gigaom.com/2011/03/18/netflix-big-data/> [Accessed 26 Jan 2017].
- Lessig, L., 2009. *Code: And other laws of cyberspace*. New York: Basic Books.
- Lohr, S., 2015. *Data-ism*. Oneworld Publications.
- Martijn, M. and Tokmetzis, 2016. *Je hebt wél iets te verbergen*. Amsterdam: De Correspondent.
- McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., and Barton, D., 2012. Big data. The management revolution. *Harvard Bus Review*, 90 (10), 61–67.
- Moerel, L. and Prins, C., 2016. *Privacy for the homo digitalis: Proposal for a new regulatory framework for data protection in the light of Big Data and the internet of things*. Tilburg: Tilburg Institute for Law, Technology, and Society.
- Moor, J., 2006. The Dartmouth College artificial intelligence conference: The next fifty years. *Ai Magazine*, 27 (4), 87.
- Muehlhauser, L. and Salamon, A., 2012. Intelligence explosion: Evidence and import. *In: Singularity Hypotheses*. Springer, 15–42.
- Müller, V.C. and Bostrom, N., 2016. Future progress in artificial intelligence: A survey of expert opinion. *In: Fundamental issues of artificial intelligence*. Springer, 553–570.
- National Science and Technology Council, 2016. *Preparing for the Future of Artificial Intelligence*. Washington: The White House.
- Pariser, E., 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.
- Pasquale, F., 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Postma, R., 2016. 'Er is een nieuw verhaal nodig om mensen te prikkelen'. *nrc.nl*.
- Richards, N.M. and King, J.H., 2013. Three paradoxes of big data. *Stan. L. Rev. Online*, 66, 41.
- Ruthkin, A., 2016. People will follow a robot in an emergency – even if it's wrong. *New Scientist*, 16 Feb.
- Sadeh, N., Acquisti, A., Breaux, T.D., Cranor, L.F., McDonald, A.M., Reidenberg, J., Smith, N.A., Liu, F., Russell, N.C., and Schaub, F., 2014. Towards Usable Privacy Policies: Semi-automatically Extracting Data Practices From Websites' Privacy Policies. SOUPS.
- Schoemaker, P., 1995. *Scenario Planning: A Tool for Strategic Thinking*. MIT Sloan Management Review.

- Schwab, K., 2017. *The fourth industrial revolution*. Penguin UK.
- Senate, U.S., 2013. A Review of the Data Broker Industry: Collection, use, and sale of consumer data for marketing purposes. *Washington, DC: Committee on Commerce, Science, and Transportation, US Senate*.
- Shultz, D., 2015, and Am, 10:30, 2015. Internet search engines may be influencing elections [online]. *Science / AAAS*. [Accessed 26 Jan 2017].
- Singer, N., 2012. Acxiom, the Quiet Giant of Consumer Database Marketing. *The New York Times*, 16 Jun.
- Snijders, D., 2016. *Nationale Toekomst Monitor 2016. Hoe kijken Nederlanders naar technologie en de toekomst?* Den Haag: STT.
- Solove, D.J., 2006. A taxonomy of privacy. *University of Pennsylvania law review*, 477–564.
- Tanz, J., 2017. Soon We Won't Program Computers. We'll Train Them Like Dogs [online]. *WIRED*. [Accessed 26 Jan 2017].
- Times, S.T.T.N.Y., 1958. New navy device learns by doing. *The New York Times*, 8 Jul.
- Timmer, J., Elias, I., Kool, L., and van Est, R., 2015. *Berekende risico's: Verzekeren in de datagedreven samenleving*. Den Haag: Rathenau Instituut.
- Turkle, S., 2011. *Alone together: Why we expect more from technology and less from ourselves*. New York: Basic Books.
- Van Engelen, M., 2015. 'Computers zijn nog net zo dom als dertig jaar geleden'. *Vrij Nederland*.
- Van Verschuer, N., 2014. Liefdeloosheid in kapitalistische tijden. *Vrij Nederland*, 5 Aug.
- Visscher, M., 2016. 'Ons oordeel is belangrijker dan Big Data'. *Trouw*.
- Wiener, N., 1999. Some moral and technical consequences of automation. *Resonance*, 4 (1), 80–88.
- Working Group on Data Protection in Telecommunications, 2014. *Working Paper on Big Data and Privacy. Privacy principles under pressure in the age of Big Data analytics*.
- Yudkowsky, E., 2008. Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks*, 1 (303), 184.

De auteur

Dr. Dhoya Snijders werkt sinds 2015 als onderzoeker en projectleider voor STT. Hij heeft een Mastertitel in de Wijsbegeerte van de Universiteit van Amsterdam, in het Publiek Informatie Management van de Erasmus Universiteit en in de Organisationswetenschappen van de Vrije Universiteit. Aan de Vrije Universiteit promoveerde hij op een onderzoek naar overheidsclassificaties en hun sociale gevolgen. Tot zijn indiensttreding bij STT werkte hij als adviseur op het snijvlak van ICT en beleid aan nationale en internationale overheidsprojecten. Momenteel voert hij voor STT onderzoek uit naar de rol die technologie kan hebben op de manier waarop we leren.



A photograph of a bright blue sky filled with numerous small, white, fluffy clouds. The clouds are scattered across the frame, with some larger, more prominent ones in the upper right and lower left corners. The overall scene is bright and clear.

data is macht

Weet u zeker dat deze tekst geschreven is door een mens? Zou het kunnen zijn dat de tekst geschreven is door kunstmatige intelligentie? Net als dat de illustratie op de kaft ontworpen is door een zelflerend systeem is één van de co-auteurs van deze publicatie een getraind neurale netwerk. Zou het deze tekst kunnen zijn?

Er is steeds meer data en datasystemen worden steeds intelligenter. Hoewel de technologische ontwikkelingen doorrazen en bedrijven hun diensten en toepassingen zo snel mogelijk op de markt proberen te krijgen, is er weinig maatschappelijke discussie over de sociale gevolgen en kansen van deze technologie. Deze STT-publicatie verkent aan de hand van toekomstbeelden en scenario's de impact van Big Data op onze samenleving en dient ter inspiratie en provocatie.

De volgende auteurs schetsen in expertbijdragen de belangrijkste data-vraagstukken voor de toekomst: Dirk Helbing (ETH Zurich), Marijn Janssen (TU Delft), Djeevan Schiferli (Innogy), Maurits Kreijveld (publicist), Jeroen van den Hoven (TU Delft), Mireille Hildebrandt (Radboud Universiteit), Eric van Tol (Fontys), Rogier Creemers (Universiteit Leiden), Maurits Martijn (De Correspondent), Jeroen Dijkhoorn (SAS), Brenno de Winter (publicist), Theo de Vries (Universiteit Twente) en Emma AI (een kunstmatig intelligent systeem).

**Stichting
Toekomstbeeld
der Techniek**

www.stt.nl



ISBN 978-94-91397-15-8



9 789491 397158 >