6 Algoritmes temmen zonder overspannen verwachtingen. Een nieuwe uitdaging op de bestuurstafel

Sander Klous en Frank van Praat

1. Inleiding

Nu we steeds meer beslissingen baseren op algoritmes (al dan niet op basis van kunstmatige intelligentie) komen daarmee ook voor bestuurders nieuwe strategische thema's op. Welke risico's loopt hun organisatie als de kwaliteit of betrouwbaarheid van deze algoritmes onvoldoende blijkt te zijn? Hoe beheersen ze deze risico's? Hoe kunnen ze aantoonbaar laten zien dat de maatschappij deze algoritmes kan vertrouwen? En hoe zorgen ze dat daar geen overspannen verwachtingen over gaan leven?

Op 1 augustus 2018 laten agenten een groepje Amerikaanse toeristen stoppen die zojuist door de Piet Hein tunnel zijn gefietst. Het gaat om een tunnel zonder fietspad waar de auto's je met snelheden van rond de 100 km voorbijrazen. Stadszender AT5 was erbij met een camera en het filmpje op YouTube laat drie dingen zien.[1] Ten eerste dat de toeristen de tunnel in gestuurd werden door hun navigatie-appje: het onderliggende algoritme dirigeerde hen een tunnel in waar ze met de fiets niets te zoeken hadden. Ten tweede dat ze een grenzeloos vertrouwen hebben in technologie en de aanwijzingen slaafs opvolgden zonder zelf na te denken. Want de agent die erop wees dat de Amerikanen toch zelf ook konden zien dat deze tunnel off limits was voor fietsers kon alleen maar rekenen op onbegrip. Ten derde dat het altijd de schuld van iemand anders is. Want een van de toeristen kondigde op hoge toon aan Apple juridisch te gaan vervolgen als de politie hen op de bon slingert. Hem treft geen blaam.

Precies deze drie zaken zijn exemplarisch voor de invloed van kunstmatige intelligentie – of specifieker: algoritmes – op de maatschappij in bredere zin: (1) Onze beslissingen worden meer en meer gestuurd op basis van data en de algoritmes die deze data gebruiken. (2) We gebruiken de technologie op slaafse wijze en vertrouwen het blindelings. (3) Als het mis gaat wordt er zo snel mogelijk een schuldige gezocht.

1 https://www.at5.nl/artikelen/184990/apple-navigatie-stuurt-fietsende-toeristen-door-de-piet-heintunnel.

De vraag komt dan ook steeds vaker op of de algoritmes onder de motorkap wel deugen. Onder andere vanwege de reactie van de Amerikaanse toerist – 'het is de schuld van Apple' – zou elke bestuurder of commissaris daar eens goed over na moeten denken en inventariseren welke risico's de organisatie op dit vlak loopt. Tien tegen één dat dan ook blijkt dat een organisatie heel veel geld uitgeeft aan het beheersen en controleren van financiële processen (al dan niet om te voldoen aan wet- en regelgeving) en dat er nauwelijks middelen worden vrijgemaakt om in control te zijn en blijven over de algoritmes (en daarover helder te communiceren). De risico's daarvan zijn echt niet alleen maar relevant voor wie een zelfrijdende auto op de markt brengt, maar ook voor banken, verzekeraars, vliegmaatschappijen en tal van andere organisaties.

Het is dan ook tijd voor een strategische herbezinning over hoe om te gaan met deze risico's en, minstens even belangrijk, hoe de verwachtingen op dit punt te managen. Dat is het onderwerp van analyse in dit hoofdstuk.

2. Hoge maatschappelijke eisen

Dat begint met een analyse van maatschappelijke ontwikkelingen en verwachtingen. Wie de mediaberichten een beetje volgt lijkt eigenlijk maar één conclusie te kunnen trekken: er gaat heel veel mis bij bedrijven en overheden. Frauderende topbestuurders, falend toezicht op uitkeringsfraude, bankiers die teveel risico's nemen, technologiebedrijven die persoonlijke data misbruiken en overheidsinstellingen die zo lek als een mandje blijken te zijn. De reactie op de nieuwsberichten lijkt ook al jaren een voorspelbaar patroon te volgen: er komt een roep om meer regels, meer controles en meer transparantie om daarmee het onder druk staande vertrouwen weer te herwinnen. Het kalf is verdronken, en de put moet in elk geval zo snel mogelijk worden gedempt. Halve maatregelen zijn niet gewenst.

Er lijkt dan ook sprake te zijn van een hardnekkige vertrouwenscrisis en een maatschappij met zero-tolerance ten opzichte van fouten. Maar is dat ook daadwerkelijk zo? Hierna doen we drie observaties (en/of nuanceringen) daarover, die van groot belang zijn om de verwachtingen goed te managen, ook ten aanzien van (een beheerst gebruik van) algoritmes.

De eerste observatie: er gaat heel veel goed, waarschijnlijk meer dan we denken.
Wie zich een goed beeld van de werkelijkheid wil vormen, kan maar beter geen kranten en andere media lezen. Dat is kortgezegd de visie van de bekende hoogleraar psychologie Steven Pinker. Gebaseerd op een stevige reeks aan feiten laat hij in zijn boek Enlightenment Now[2] zien dat het best goed gaat met de wereld. Maar uit een stroom mediaberichten over terrorisme, hongersnoden, natuurrampen

2 Enlightenment Now, 2018, Allen Lane.

en financiële crises doemt een heel ander beeld op: het beeld dat onze beschaving aan de rand van de afgrond staat. Er spelen grofweg twee redenen een rol bij dat verschil: Ten eerste gaat nieuws vrijwel altijd over incidenten en niet over trends; en ten tweede hebben we als mens wat systeemfouten – cognitieve vooroordelen – die onze blik vertroebelen. Dat lijkt ook in onze perceptie op dit domein een rol te spelen: negatieve berichten over bedrijven die slordig, onkundig of slechtwillend met data en/of algoritmes omgaan krijgen in de media veel aandacht. De vooruitgang onttrekt zich juist aan ons blikveld.

De tweede observatie: er is een leiderschapscrisis, geen vertrouwenscrisis

Ten tweede is het maar de vraag of er wel echt sprake is van een vertrouwenscrisis. Robert Phillips bepleit in zijn boek Trust Me, PR is Dead[3] dat het niet zozeer gaat om een vertrouwenscrisis maar om een leiderschapscrisis. Leiders moeten vooral met acties het vertrouwen herwinnen. Phillips beschrijft onder meer hoe een CEO te maken had met een grote brand in zijn toeleveringsketen met 100 dodelijke slachtoffers. Een ramp in humanitaire zin. Maar mogelijk ook een ramp voor de reputatie van het bedrijf. De bezorgde CEO belt zijn adviseur Phillips en zegt dat zijn voorzitter van de Raad van Commissarissen hem op de huid zit. Want: 'we're failing to get our message across. We are not emphasising our CSR [corporate social responsibility] credentials well enough.' De CEO weet niet goed wat te doen en vraagt zijn adviseur Phillips om goede raad. Zijn antwoord: 'You start with actions, not words.' Het voorval legt anekdotisch bloot dat het meer om acties dan om mooie woorden gaat. Dat geldt niet alleen voor wie te maken heeft met brand maar ook voor wie te maken heeft met incidenten rondom data en/of algoritmes. Als leiders laten zien dat hun organisatie betrouwbaar is, dan komt het met dat vertrouwen vanzelf goed. Heel simpel: alleen wie de goede dingen doet zal vertrouwen oogsten.

De derde observatie: de maatschappij stelt hoge(re) ethische eisen

Ten derde stelt de maatschappij hoge verwachtingen aan de mores van bedrijven en overheden. De tijdgeest is echt anders dan enkele decennia geleden. Een mooie analogie is zichtbaar in een discussie die in 1982 speelde. Dit draaide rondom een geautomatiseerd reserveringssysteem voor commerciële vluchten geïntroduceerd, genaamd SABRE, oftewel Semi- Automated Booking and Reservation Environment. Aanvankelijk werd dit systeem ontwikkeld voor American Airlines, maar later werden er ook vluchten van andere maatschappijen aan toegevoegd. Gaandeweg werd echter steeds duidelijker dat het systeem het klantbelang van de reiziger niet altijd voorop stelde: de eigen vluchten stonden vaak bovenaan, ook als deze duidelijk duurder waren en qua reisschema minder goed aansloten op de reiswensen. Talrijke klachten van onder meer reisagenten leidden uiteindelijk onder meer tot een hoorzitting voor het Amerikaanse congres in 1982. Robert Crandall, destijds CEO van American, was daar opvallend eerlijk in zijn betoog: "The preferential display of our flights, and the corresponding increase in our market

3 Trust Me, PR is Dead, Cornerstone, 2015.

share, is the competitive raison d'etre for having created the [SABRE] system in the first place." Business as usual dus in een bedrijfstak die door Wall Street Journal Journalist Thomas Petzinger destijds in zijn boek Hard Landing[4] – de bron van bovenstaand citaat – werd betiteld als 'a nasty, rotten business.'

37 jaar later hebben techreuzen als Apple en Google het aan de stok met autoriteiten over hoe ze al dan niet hun macht misbruiken, onder meer door onderhuids de eigen producten en diensten prominenter te laten zien. Marc Zuckerberg moet voor het Amerikaanse congres getuigen en vliegt ook naar Europa om uitleg te geven over hoe Facebook met data omgaat. Op tal van fronten spelen issues die eigenlijk vergelijkbaar zijn met de casus rondom SABRE van lang geleden. Neem bijvoorbeeld webshops die op basis van algoritmes persoonlijke tips geven over wat je misschien nog meer leuk vindt en waar je de beste deal daarvoor kunt vinden. Deugen die aanbevelingen wel? Een onderzoek in 2016 suggereerde dat Amazon niet de beste opties voor de klant bood maar vooral voor eigen (commerciële) parochie preekte.[5]

Er is echter een groot verschil ten opzichte van vroeger: het is nu ondenkbaar dat Mark Zuckerberg of andere topbestuurders net als Robert Crandall destijds zouden zeggen dat bevooroordeelde systemen simpelweg hun *raison d'etre* zijn. Een golf aan hoon zou hun deel zijn en de (reputatie)schade zou niet te overzien zijn. Wat wel en niet mag is dus mede afhankelijk van hoe de normen zich maatschappelijk ontwikkelen. Dat geldt niet alleen ten aanzien van het gebruik van data en algoritmes, maar ook ten aanzien van zaken als duurzaamheid, mensenrechten en andere aspecten.

3. Het gebruik van algoritmes levert nieuwe strategische risico's op

Met die achtergrond over de maatschappelijke ontwikkelingen is het zaak om te analyseren welke risico's er kleven aan het gebruik van algoritmes. Zoals we in de inleiding van dit hoofdstuk al lieten zien met het voorbeeld van de fietsers die hun navigatie-app al te blindelings vertrouwen is duidelijk dat algoritmes steeds meer invloed krijgen op ons leven. De geautomatiseerde modellen helpen ons op een breed spectrum beslissingen te nemen op basis van de informatie die we erin stoppen. Huisartsen worden door algoritmes bijvoorbeeld ondersteund bij het stellen van diagnoses. Nieuwsmedia schotelen ons een gepersonaliseerde newsfeed voor op basis van wat ze van ons weten. De politie voorspelt op basis van algoritmes in welke straten surveillance het meest zin heeft. En zo kunnen we nog wel even doorgaan.

4 Hard Landing, Three Rivers Press, 1995.
5 https://www.technologyreview.com/s/602442/amazons-algorithms-dont-find-you-the-best-deals/.

Het oude Orwelliaanse motto was dat *Big Brother is watching you*. De realiteit is echter dat we steeds meer – zichtbaar en onzichtbaar – gestuurd worden in ons handelen door systemen: *Big Brother is guiding you*.[6] Het systeem – het algoritme – komt steeds meer in de driver's seat te zitten bij het nemen van beslissingen. Het algoritme wordt langzamerhand dan ook steeds meer bepalend waar het gaat om bijvoorbeeld nieuw overheidsbeleid, keuzes over marktintroducties en afwegingen over fusies en overnames. Een mooi *sign of the times*: naar verluidt[7] zijn er techbedrijven die het budgetteringsproces geheel in handen geven van een algoritme.

Een risicoverhogende factor daarbij is dat we als mens tamelijk slaafs de adviezen van een computer opvolgen. Dat heeft te maken met de zogeheten *automation bias*. In geautomatiseerde omgevingen blijken we onwrikbaar geloof te hebben in technologie, variërend van het voor zoete koek aannemen van suggesties van een spell checker tot een piloot die blindvaart op de technologie in zijn cockpit. Ter illustratie en als mogelijke verklaring voor het fietsincident in de Piet Hein tunnel: een onderzoek wees uit dat we bepaalde delen van ons brein 'uitzetten' als we gebruik maken van de aanwijzingen van een navigatiesysteem.[8]

Nu ons lot steeds meer in handen komt van technologie – de algoritmes met onderliggende modellen – komt ook de noodzaak op erop toe te zien dat we die modellen niet voeden met misinformatie, dat de modellen zelf deugen en dat het gebruik door mensen op verantwoorde wijze gebeurt. Het gaat hier zowel om eenvoudige algoritmes – rekenregels in systemen – als om complexere lerende algoritmes – gebaseerd op kunstmatige intelligentie.

Hierbij is sprake van een belangrijk strategisch risico en we zien de afgelopen jaren de nodige 'ongevallen' gebeuren die dat risico onderstrepen. Daarbij gaat het niet alleen om internetbedrijven die onverantwoord met data omgaan, maar ook om bedrijven in andere sectoren. Een voorbeeld is het aan de grond houden van de Boeing 737 Max na twee ongevallen. Uit eerste analyses bleek dat het mogelijk gaat om kunstmatige intelligentie aan boord van het vliegtuig die verkeerde conclusies trok op basis van sensordata: een gevalletje 'confused AI'.[9]

Dichterbij huis is de witwasaffaire van ING – met als gevolg een boete van bijna 800 miljoen euro – in wat meer abstracte zin ook een algoritmeprobleem. De parameters van de systemen bleken niet in staat om een relevante en te behappen hoeveelheid 'red flags' uit de data te filteren. Het gevolg: een overload aan meldingen met veel 'false positives' waar onvoldoende menskracht beschikbaar was om er effectief mee om te gaan. Ook het impactvolle dieselschandaal van

6 Vertrouwen in de slimme samenleving, 2017, Klous en Wielaard.
7 New Kid zoekt Accountant, Accountant Q1 2019.
8 https://www.theguardian.com/science/2017/mar/21/all-mapped-out-using-satnav-switches-off-parts-of-the-brain-study-suggests-navigation.
9 https://www.news.com.au/technology/innovation/inventions/how-a-confused-ai-may-have-fought-pilots-attempting-to-save-boeing-737-max8s/news-story/bf0d102f699905e5aa8d1f6d65f4c27e.

Volkswagen is in de kern het gevolg van ondeugdelijke algoritmes in de emissietest. Een laatste voorbeeld in dit rijtje: er komt steeds meer twijfel over hoe de wereld van internetadvertenties 'onder de motorkap' werkt. De belofte van online adverteren is dat de effecten van reclame steeds beter zijn uit te drukken in cijfers. Met een verwijzing naar de populaire Netflix serie rondom reclameman Don Draper: 'Mad men' maakt plaats voor 'Math men'. Maar de klanten van die Math men vertrouwen het steeds minder.[10]

4. Nieuwe risico's vragen om nieuwe afwegingen over beheersing

Bestuurders hebben een verantwoordelijkheid voor de lange termijn continuïteit van een onderneming en moeten vanuit die verantwoordelijkheid investeren in beheersing van risico's. Dat is niet alleen vanuit defensief oogpunt nodig, maar ook (of vooral) omdat beheerst omgaan met risico's strategische voordelen oplevert. Wie beter inzicht heeft in risico's, kan betere beslissingen nemen op dit vlak en heeft daarmee een troef in de concurrentiestrijd. Op financieel vlak is het instrumentarium daarvoor al volwassen. Er worden middelen vrijgemaakt voor interne beheersing en -controle, risicomanagement en accountantscontrole, vanuit de wetenschap dat ondeugdelijk financieel beheer grote risico's oplevert.

In het geval van ondeugdelijke algoritmes zijn die risico's ook groot en in een aantal gevallen waarschijnlijk zelfs groter dan de financiële risico's. Ervaringen in de praktijk laten zien dat veel organisaties op dit vlak nog geen volwassen benadering hebben voor het beheersen van risico's. Sterker nog: in veel gevallen is er nog niet eens een analyse gemaakt of het nodig is hier middelen voor vrij te maken.

Met de verdere opkomst van algoritmes – en de groeiende impact ervan – is dat niet langer houdbaar. Niets doen is natuurlijk een keuze, maar dan accepteert een bestuurder ook welbewust dat er risico's zoals hiervoor beschreven kunnen optreden. In de meeste gevallen zal het niet de beste keuze zijn en is het zinvoller om het thema 'in control over AI' nadrukkelijk in de boardroom te analyseren en er consequenties aan te verbinden.

De million-dollar-question is: hoe kunnen organisaties komen tot een volwassen beheersing van de risico's die voortvloeien uit de toepassing van algoritmes?

Puur methodologisch gezien is er eigenlijk weinig verschil met het beheersen van risico's op andere domeinen. Er zijn drie categorieën risico's te onderscheiden. Achtereenvolgens gaat het om operationele risico's (een ondeugdelijk algoritme), juridische risico's (een algoritme dat zich niet aan de wet houdt), en reputatierisico's (een algoritme dat dingen doet die niet passen bij wat stakeholders ervan

10 https://decorrespondent.nl/9090/dit-is-de-nieuwe-internetbubbel-online-advertenties/5468662079l0-cee4cb5f.

verwachten). Deze drie risico's monden uit in een financieel risico. Precies deze lijn van denken over risico's is ook toepasbaar op andere terreinen. Een financiële instelling die een nieuw product in de markt zet, zal immers ook de behoefte hebben om in control te zijn over de drie genoemde risicocategorieën.

Ook op een ander punt is er conceptueel weinig nieuws onder de zon: het beheersen van risico's is een uitvloeisel van een analyse over twee assen. Op de ene as staat de kans dat een gebeurtenis zich voordoet. Op de andere as de impact.

Impact

Ten aanzien van dat laatste – de impact – is de uitdaging voor een goede inschatting waarschijnlijk het grootst, simpelweg omdat er weinig ervaringen – en dus weinig data – beschikbaar zijn. Het gaat om terra incognita als gevolg van het nog jonge karakter. Het doet wat dat betreft denken aan hoe het risicobeheer van cyber zich de afgelopen tien jaar ontwikkelde. Ook voor cybercrime was (en is) vaak niet duidelijk wat de impact van een incident nu echt is. Voor verzekeraars vormde dat ook een uitdaging: als zij geen data hebben om hun modellen te vullen, kunnen ze geen producten ervoor in de markt zetten, of kunnen dat alleen maar tegen heel hoge kosten doen.[11] Een min of meer identieke situatie is er nu ten aanzien van de risico's van algoritmes: we zien de eerste 'ongevallen' plaatsvinden maar hebben nog weinig inzicht in wat de impact daarvan is.

Kans

Ten aanzien van de hiervoor genoemde eerste as – de kans dat een gebeurtenis zich voordoet – is de dynamiek voor een deel wel heel anders dan het geval is bij cyber. Wat vergelijkbaar is: de grote onzekerheid en complexiteit. Net zoals cyberincidenten uit allerlei hoeken en gaten kunnen opdoemen – in jargon: er is sprake van een groot 'threat surface' – kunnen ook incidenten rondom algoritmes op heel veel verschillende domeinen plaats vinden. Er is echter wel een verschil in de 'beheersbaarheid' van de kans. In het geval van algoritmes is het mogelijk om de systemen te 'tweaken' en daarmee exact in te regelen hoeveel kans er is op *false positives* of *false negatives,* een van de zaken die mis kunnen gaan. Ook toezicht en (kwaliteits-) controle op de werking van algoritmes is goed mogelijk. Bij cyberrisico's bestaan daartoe veel minder opties.

Kortom

De (traditionele) modelmatige aanpak van risicobeheersing is dus ook van toepassing op dit relatief nieuwe domein. Maar deze modellen zijn pas waardevol als er voldoende bewustwording is. Dat is dan ook een belangrijk thema. Mede daarom is het zaak om de verantwoordelijkheid ervoor 'in te bouwen' waar deze thuishoort. De controlelaag bovenop de business moet beperkt blijven, en de business zelf moet juist verantwoordelijkheid nemen op dit aspect. Dit is niet alleen

11 Een ontwikkeling die hier in 2015 werd beschreven: https://riskandinsurance.com/cyber-challenges-still-evolving/ .

de beste oplossing om de cost of control overzichtelijk te houden, maar ook om het ondernemerschap de ruimte te geven.

Dat ondernemerschap vraagt om dynamiek en wendbaarheid. Ondernemen gaat per definitie niet zonder risico's en dat geldt ook ten aanzien van het ontwikkelen van algoritmes – en daarmee het doorvoeren van innovaties in producten, diensten en bedrijfsmodellen. Dat betekent ook dat het onwenselijk is om een te strak beheersingsinstrumentarium te gebruiken want daarmee gaan waarschijnlijk strategische kansen verloren omdat dit de creativiteit en flexibiliteit beperkt.

5. Waterdichte garanties zijn er niet

Iedereen weet dat een foutloze maatschappij een illusie is. Fouten maken hoort gewoon bij het leven. Streven naar 100% zekerheid over de werking van technologie is niet alleen onhaalbaar maar ook onwenselijk. Dat wekt immers de verkeerde verwachtingen op en legt daarmee potentieel een bommetje onder het vertrouwen als het een keer mis gaat.

Ook in dat kader kunnen we veel leren van hoe het thema cyber security zich heeft ontwikkeld. Rondom cyber security hing lange tijd de teneur dat we moesten streven naar 100% zekerheid. Onze data mocht onder geen beding op straat komen te liggen en de beveiliging moest dat garanderen. Het leidde tot steedse verdergaande beveiligingsmaatregelen. Inmiddels draait de wind en beseft men steeds beter dat 100% zekerheid niet haalbaar is en dat pogingen om dat te doen waarschijnlijk te kostbaar en gebruiksonvriendelijk zijn. Veel organisaties realiseren zich dat ze gehackt zullen worden of dat misschien ongemerkt al zijn. Het is business as usual geworden. Precies daarom focussen ze zich met gezond boerenverstand op het beheersen van de risico's. Het 'slot' hoeft niet onbreekbaar te zijn, als het maar is afgestemd op de waarde van de spullen die binnen liggen. Eigenlijk net zoals een juwelier bewust keuzes maakt over inbraakbeveiliging en accepteert dat daar een bepaald risico bij hoort.

Net zo min als bij cyber security het geval is, is er ook geen absolute zekerheid mogelijk over de werking van algoritmes. Zelfrijdende auto's zullen soms crashen en wervingsalgoritmes zullen kandidaten niet altijd eerlijk behandelen. Het is zaak om de verwachtingen daarover goed te managen, zo leren de ervaringen vanuit cyber security ons.

Die ervaringen bieden ook nog andere aangrijpingspunten om te leren. Een goede aanpak voor cybersecurity bouwt voort op een analyse van de eigen activiteiten en het onderscheiden van daaraan verbonden risico's in verschillende categorieën. Een dergelijke compartimentering helpt om de impact van bepaalde incidenten in kaart te brengen, de spreekwoordelijke kroonjuwelen te identificeren en de maatregelen daarop af te stemmen. Simpel gesteld: de impact van een gehackt mailaccount van

een receptioniste is van een andere orde dan malware die zich nestelt in de systemen van een kerncentrale. Een dergelijke compartimentering is ook wenselijk voor de risico's van ondeugdelijke algoritmes. Als de doorontwikkeling van een algoritme er een paar dagen voor zorgt dat er in webshop geen goede aanbevelingen worden gedaan aan klanten, dan is dat vervelend maar zijn de gevolgen wel te overzien. Maar als een vliegtuig neerstort omdat de AI onder de motorkap niet goed functioneerde dan is dat niet alleen vanuit menselijk oogpunt rampzalig maar ook een serieuze bedreiging voor het bestaansrecht van de vliegtuigfabrikant.

6. Nadenken over ethiek

Naast een analyse van de potentiële gevaren – die elke bestuurder aan het denken zouden moeten zetten over de niet geringe risico's van ondeugdelijk functionerende algoritmes – is er ook nog een belangrijke ethische kant aan. Aandacht daarvoor is minstens zo belangrijk om de verwachtingen goed te managen.

Dat heeft alles te maken met het feit dat we in 'the age of the algorithm' een spannende nieuwe optie krijgen: We kunnen ethiek programmeren. 'Code is Law', zo schreef Lawrence Lessig in 1999 al in zijn boek[12] waarin hij beschreef dat programmeurs een sterk sturende rol hebben in ons leven door de systemen die ze bouwen. In die tijd speelde kunstmatige intelligentie nog niet zo'n sterke rol in onze samenleving en ging die sturende werking van softwarecode maar beperkt over ethiek. Dat is nu heel anders. Data scientists stoppen – al dan niet bewust – ethiek in het algoritme dat ze ontwikkelen voor bepaalde taken. Ze zijn dan ook de nieuwe hogepriesters van de maatschappij geworden, maar zijn juist helemaal niet opgeleid op het vlak van ethiek. Bovendien: er zijn eigenlijk nog vrijwel geen normen op dit vlak – al is wel duidelijk uit de eerdere analyse over SABRE dat de lat steeds hoger wordt gelegd en dat er veel van bedrijven wordt verwacht.

Maatschappelijk hebben we in elk geval nog geen conclusie getrokken over wat de norm moet zijn. Daarbij spelen overigens twee vragen een rol. De eerste vraag is welke activiteiten / beslissingen we bereid zijn om in handen van algoritmes te leggen. De tweede is welke ethische voorwaarden we moeten programmeren als we algoritmes gebruiken. Voor beide vraagstukken geldt dat dit nog grotendeels terra incognita is.

Het belang van die constatering neemt nog verder toe als we goed beseffen hoeveel onvoorziene impact algoritmes kunnen hebben op de samenleving. Technologie die belooft een bepaald probleem op te lossen kan op een ander vlak juist weer een probleem creëren. Een mooi voorbeeld is hoe veel steden honderd jaar geleden een probleem hadden: een overdaad aan paardenpoep in de straten. De komst van de

12 Code and Other Laws of Cyberspace, Basic Books, 1999.

auto werd gezien als een welkome oplossing. Geen beleidsmaker dacht na over de nadelen – zoals vervuiling, lawaai, congestie.

Die nadelen kennen we inmiddels. Congestie kunnen we eigenlijk zien als de paardenpoep van de 21e eeuw. En er zijn natuurlijk ook andere problemen zoals een gebrekkige luchtkwaliteit. Opnieuw lijkt nieuwe technologie de oplossing voor deze problemen en dit keer is dat de Smart City, een concept waarbij we op basis van data nieuwe mogelijkheden krijgen om te sturen en organiseren. Tal van steden zetten experimenten op rondom Smart Cities. En bij nieuwe stedelijke gebieden reiken de ideeën een stuk verder, zoals onder meer te zien is bij de plannen voor de Toronto Waterfront area, waar Google dochter Sidewalk Labs een groot project heeft gewonnen. Belangrijke onderwerpen van veel plannen zijn logistiek en mobiliteit. Juist op die domeinen ontstaan door nieuwe technologie andere sturingsmogelijkheden.

De uitdaging is om dit keer wel voorbij de paardenpoep te kijken. De voordelen van een datagedreven aanpak in een stedelijk gebied zijn evident. Als een ambulance een spoedje heeft kunnen de verkeerslichten op de route op groen, al dan niet volautomatisch op basis van een algoritme. Als er tijdens een groot evenement massa's mensen in het gedrang dreigen te komen, kan daar eveneens proactief op worden ingegrepen. Als de luchtkwaliteit in een wijk verslechtert, kan een algoritme mogelijk zelfs navigatiesystemen in auto's even tijdelijk via een andere route leiden.

Maar de technologie heeft ook een duistere kant. Er is bijvoorbeeld niet veel fantasie nodig om je voor te stellen dat een ambtenaar een zelfrijdende auto van een individu straks verbiedt om een bepaalde wijk binnen te rijden. Het scenario "Computer says no" voltrekt zich dan. Hopelijk op basis van een ethisch verantwoorde afweging. Hoe bewaken we dat nieuwe mogelijkheden op een integere manier worden ingezet zonder dat daarbij groepen of individuen onterecht worden benadeeld?

Kortom: de uitdaging is dan ook om voorbij de spreekwoordelijke paardenpoep te kijken en te komen tot een beheerst gebruik van de technologie. Maar opnieuw: dat kan alleen als we een glashelder beeld hebben van de normen die we maatschappelijk stellen aan die technologie! En juist daar hangt nu nog veel mist.

De vraag is en blijft dus: Wie gaat die normen formuleren? Op dit moment lijkt niemand zich aan de hete aardappel te willen branden. Wetgevers hebben het thema nog maar net ontdekt – en lopen haast per definitie jaren achter op technologische ontwikkelingen. Sterker nog, in veel gevallen is er sprake van wat Koen Frenken, professor aan de Universiteit van Utrecht, 'reverse technology assessment'[13] noemt. De klassieke volgorde van technologiebeoordeling is wetenschappelijk

13 https://www.uu.nl/sites/default/files/20160211-uu_oratie-frenken.pdf.

onderzoek, een normatieve publieke discussie, regulering en dan marktintroductie. Zo gaat het bijvoorbeeld altijd bij nieuwe medicijnen en nieuwe typen vliegtuigen. Maar in de digitale economie gaat het vaak heel anders. Bedrijven lanceren eerst een nieuw platform of dienst, daarna volgt pas de normatieve discussie, en daarna pas het wetenschappelijk onderzoek. Ze doen dat ook vaak met een enorm groeitempo, zodat de normerende werking van technologie zich al heeft voltrokken voordat er sprake kan zijn geweest van een goede discussie.

7. Conclusie

Bestuurders zijn verantwoordelijk voor de lange termijn continuïteit van hun organisatie. Daarbij hoort dat ze risico's scherp op het netvlies hebben en gericht maatregelen nemen om die risico's te beheersen. Voor de financiële risico's en de financiële informatievoorziening is daar sprake van een volwassen benadering, gestoeld op een lange historie. Nu Artificial Intelligence/Algoritmes een grotere rol gaan spelen is het ook op dit nieuwe domein zaak dat bestuurders aandacht hebben voor de risico's en laten zien dat ze er maatregelen nemen. Precies zoals in dit hoofdstuk betoogd: leiderschap draait om daadkracht en als die daadkracht er is kan er vertrouwen ontstaan.

Nadenken over de beheersing van algoritmes lijkt voor veel bestuurders misschien nog een ver-van-mijn-bed-show. Maar niets is minder waar, vooral door de exponentiële groei van zowel rekenkracht als de hoeveelheid data zal de rol van Artificial Intelligence zeer snel groeien. Bovendien kunnen de risico's ook voortvloeien uit andere vormen van geautomatiseerde besluitvorming waar (nog) geen Artificial Intelligence aan te pas komt. Burgers beginnen steeds beter te begrijpen dat er wat op het spel staat. In een onderzoek van KPMG onder 1100 Nederlanders bleek onder meer dat een overgrote meerderheid van 80 procent een vorm van toezicht op het gebruik van algoritmes wil.[14] Het is een duidelijk signaal. Niet alleen aan de overheid, maar ook aan bestuurders van bedrijven.

Zij staan voor de taak om deze technologie beheerst te gebruiken binnen hun organisatie. Dat betekent zorgen voor deugdelijke algoritmes en het bewustzijn over het belang daarvan stimuleren. En het betekent inspelen op de (vooralsnog niet scherp geformuleerde) ethische normen.

Bij het 'temmen' van de technologie is het zaak om de verwachtingen goed te managen, zeker in een kortademige samenleving zoals geschetst in dit artikel. Een goede beheersing draagt bij aan het voorkomen van incidenten als gevolg van ondeugdelijke algoritmes, maar biedt geen zekerheden. Droomwerelden bestaan immers niet, ook als het algoritme steeds meer taken overneemt van ons.

14 https://home.kpmg/nl/nl/home/insights/2019/06/vertrouwen-van-de-nederlandse-burger-in-algoritmes.html.