

# Technologie in de tussentiid. *AI, publieke waarde en democratie* tussenruimte.

**Nederlandse  
School voor  
Openbaar  
Bestuur**

Martijn van der Steen

*In opdracht van*



provincie  
Zuid-Holland



**prof. dr. Martijn van der Steen**

is bestuurder en decaan van de Nederlandse School voor Openbaar Bestuur. Daarnaast is hij bijzonder hoogleraar Strategie en Toekomst aan de Erasmus Universiteit Rotterdam en buitengewoon hoogleraar aan de Universiteit van Stellenbosch, Zuid-Afrika. Sinds 2021 is hij lid van de Raad voor Volksgezondheid en Samenleving.

## Managementsamenvatting

AI is een disruptieve technologie die de principes van de democratische rechtsstaat in de basis raakt. We bevinden ons in de tussentijd van AI: een periode waarin AI nog te onderontwikkeld is om écht op te vertrouwen, maar te goed en te diepgaand om te negeren.

AI vereist sturing door de overheid: om de principes van de democratische rechtsstaat te borgen en te beschermen en de toepassing van AI ten behoeve van de realisatie van publieke waarde te optimaliseren.

Die sturing door de overheid vereist explicitering van hoe we ons tot AI verhouden. Geloven we in toekomstige perfectie van AI, dan is er veel te zeggen voor het overgeven van het stuur aan de technologie (silicon), zoals in een zelfrijdende auto. Is de verwachting dat AI de perfectie in een bepaald domein nooit zal bereiken, dan is altijd de mens (carbon) als bestuurder nodig. En moeten mensen (carbons) altijd 'on top' van de ontwikkeling in AI (silicons) blijven.

Dit essay laat zien dat AI-toepassingen botsen met democratische en rechtsstatelijke principes: ze vereisen sturing door de overheid op die punten. Dat niet doen betekent per definitie verlies van democratische en rechtsstatelijke kwaliteit. Hier is géén zachte middenweg mogelijk.

AI-toepassingen optimaliseren de productielogica en machtsstructuren van een bestaand systeem. Stille en zwakke belangen worden door AI-toepassingen verder verzwakt, omdat AI fungeert als megafoon voor bestaande bias. Zonder expliciete sturing en bijsturing zijn gevolgen voor zwakke belangen groter dan ooit.

Bijsturing van AI is niet moeilijk, maar wel lastig: het vereist commitment, investering en lange adem. De gevolgen van AI voor democratie, rechtsstaat en publieke waarde zijn goed in beeld te brengen door het stellen van de juiste set kritische vragen. Dit essay benoemt de te stellen vragen en biedt drie strategieën voor het omgaan met AI.

Omdat AI als eerste technologie ingrijpt in de fundamenten van overheidsbestuur moeten we het als fundamentele kwestie benaderen. De democratie en rechtsstaat zijn te belangrijk om zomaar aan silicons over te laten. Tussentijd vereist expliciete strategie van carbons.

# Inhoudsopgave

---

- 1. Leven in de tussentijd van AI → 4**  
Leeswijzer → 7
- 2. Silicons en carbons: oude wijn in nieuwe zakken, of nieuwe wijn? → 8**
- 3. AI in de democratische rechtsstaat → 23**  
Politiek, rechtsstaat en AI → 23  
Technologie ten behoeve van goed overheidsbestuur → 24  
AI in de democratische rechtsstaat: een checklist voor checks en balances → 27
- 4. AI en het realiseren van publieke waarde → 30**  
Gezocht: de publieke waarde van een bibliotheek → 30  
Publieke waarde is wat het publiek waardeert → 31  
Strategisch bewegen door de driehoek → 32  
Een rechts of linksdraaiende driehoek → 33  
Technologie voor de (politiek van de) status quo → 34  
AI in de driehoek: instrument of identiteit → 34  
AI en de realisatie van publieke waarde: de driehoek in de goede richting doorlopen → 37
- 5. Denken over de toekomst van AI in het openbaar bestuur: *meer, beter, anders?* → 38**  
Denken over toekomst: meer, beter, anders → 38  
De toekomst van de zelfrijdende auto → 38  
AI en de toekomst: meer, beter, anders → 40  
Permanente imperfectie of wachten op uiteindelijke perfectie → 44
- 6. Handelingsperspectief: strategieën voor het omgaan met AI in de tussentijd → 47**  
Strategisch omgaan met brede, diepe en onbedoelde gevolgen → 47  
Strategie voor de tussentijd → 50

# 1. Leven in de tussentijd van AI

Ineens vlieg ik voorover. De auto gaat vol in de remmen, midden op de A13 ter hoogte van Delft. Om het verkeer niet te hinderen trap ik het ‘gaspedaal’ in om het automatisch remmen te overrulen en snelheid te meerderen. Mijn hart gaat tekeer en ik zie naast me een tierende medeweggebruiker die mij inhaalt. We rijden door. Geen obstakel te bekennen. Wat heeft de Tesla nou gezien om deze noodstop te maken? En waarom gebeurt het al voor de derde dag op een rij hier?

Ik rijd sinds 2014 in een Tesla Model S. Het is wat mij betreft de beste elektrische auto ooit gebouwd en ik denk dat het de enige reden is waarom de elektrificatie van automobility écht op stoom is geraakt. Tesla was niet het gevolg van een beweging naar elektrisch rijden, maar het begin ervan. Tesla forceerde een doorbraak in de transitie naar meer duurzame mobiliteit, die tot die tijd steeds door de fossiele auto-industrie was afgewenteld. In 2012 bezocht ik het bedrijf voor het eerst. Ik raakte er onder de indruk van de droom, ambitie en vreemd genoeg de realiteitszin die de toen nog ondenkbare gedachte van een ‘elektrische auto zonder compromis voor de massa’ toen al had. Een idiote en op dat moment onmogelijke gedachte, die met voortvarendheid en stelligheid werd uitgedragen. Een elektrische auto voor de massa, in een tijdperk waarin benzine en diesel nog de vanzelfsprekende standaard waren voor elke auto. En ook niemand in die massa die op een elektrische auto zat te wachten.

Maar tijdens het bezoek in 2012 aan Silicon Valley viel iets anders mij nog méér op. We maakten met een ingenieur een proefrit in een prototype Tesla Model S. Tijdens de rit vertelde hij in een bijzin dat de werkelijke doorbraak van de Tesla Model S *niet* de elektromotor of de accu was, maar de software die de auto bestuurt. De Model S is geen auto met een computer, zo legde de man uit, het is een computer met een auto. Die computer staat bovendien niet op zichzelf, maar maakt (met de auto) deel uit van een groot netwerk van andere auto's/computers. Gezamenlijk produceren al die verbonden delen van het netwerk, met elke meter die ze rijden, data waarmee de software voor de *besturing* van de auto wordt verbeterd. *Besturing?! De Model S, zo legde de ingenieur uit, is ontworpen als een zelfrijdende auto. Geen auto met software om de besturing door de bestuurder te veraangename*n, maar een auto met software om zichzelf te besturen. En met de daarvoor benodigde hardware, al is daar sinds die ochtend in 2012 nog heel veel in veranderd. De Model S heeft de elektrificatie van automobility in gang gezet, wat een historische prestatie op zich is. Maar in de geschiedenis zal Tesla – zo is mijn

overtuiging – uiteindelijk herinnerd worden als het bedrijf dat zelfrijdend vervoer mogelijk maakte. En, dát moment, zo is mijn volgende overtuiging, zullen we ons straks herinneren als het moment waarop AI het niveau heeft bereikt waarop het zelfstandig kan werken en denken in complexe veranderlijke sociale situaties. Geen tijdperk van verandering, maar een verandering van tijdperk, zou Jan Rotmans zeggen: in dit geval, *het tijdperk van artificiële intelligentie*.

Een verandering van tijdperk betekent dat we de geschiedenis kunnen indelen in de tijd vóór en de tijd ná artificiële intelligentie. Maar, zoals vaak met tijdperken en verandering, die scheidslijn is niet zo scherp te trekken. Wat is dan het tijdperk waarin we nu verkeren? Nu AI er nog niet écht is én ondertussen ook al wel een belangrijk deel van ons leven uitmaakt? We leven wat AI betreft in de **tussentijd**: *het tijdperk waarin AI er tegelijkertijd wel én niet is*. Het is een tijd waarin de Tesla moeiteloos 1400 kilometer naar Italië rijdt, zonder hapering of angstaanjagende momenten, zoals ik die van mijn eerdere autovakanties (waarin ik zelf aan het stuur zat) herinner. Een tijd waarin de Tesla zichzelf in de meest minuscule parkeerplekjes draait en kalm en klinisch elke centimeter ruimte benut en altijd op precies het goede moment in exact de ideale hoek instuurt. Waar de auto naadloos invoegt op de snelweg en nooit iets in de dode hoek mist of per ongeluk iemand afsnijdt. De Tesla is bovendien een bestuurder die nooit dronken is en die nooit ruzie met de achterbank of de bijrijder maakt. Maar, in diezelfde tijd is de auto tegelijkertijd niet in staat om een simpele rotonde te nemen, een handeling die zelfs stomdronken of driftig ruziënde bestuurders doorgaans lukt. En hij trapt soms op willekeurige plekken vol de rem in en heeft de grootste moeite met voor ‘ons’ als bestuurder heel simpele handelingen. Zoals het inhalen van een voetganger of fietser, die half op de rijbaan loopt of fietst. Tesla kan dat niet: of hij rijdt de fietser of voetganger aan, óf hij rijdt er op veilige afstand stapvoets achteraan. De *tussentijd* van AI dus: extreem goed in pakweg 99,9% van de situaties, dodelijk slecht in de 0,1%. Te goed om te negeren, te slecht om op te vertrouwen.

Dit essay gaat over het **leven in de tussentijd van AI**, waarin het te goed is om te negeren, maar te slecht om op te vertrouwen. Wat betekent bestuur in die tussentijd en waar bewegen we dan naartoe? Een cruciale vraag daarbij zal zijn wat precies de intelligentie is die AI kan of zal bereiken. En dus ook of AI het hoogste niveau van die intelligentie ooit zal bereiken. Zoals ook voor Tesla ongewis is of de auto ooit op het zogeheten ‘niveau 5’ (het werkelijk zelfrijdende niveau) zal rijden. Voor mij is de *Autopilot* van Tesla daarom dé metafoor voor het bestuderen van AI en de werking van die technologie in het bestuur en in de democratie: niet vanwege het democratisch gehalte van Tesla – dat is belabberd – maar vanwege de steile maar tevens incomplete ontwikkeling van AI. Net als in de Tesla kan AI al heel veel overnemen van de bestuurder.

Maar, nog niet alles. Het verhaal begint niet voor niets met een schijnbaar willekeurige noodstop – op de derde achtereenvolgende dag op ongeveer dezelfde plaats. *Niet* willekeurig dus, maar zeker niet accuraat. En enkele dagen later opgelost met een software-update. Autopilot is voor mij een metafoor voor de ontwikkeling van AI, op twee tegelijkertijd geheel tegenstrijdige manieren. Enerzijds laat het zien hoe ongelofelijk geavanceerd de technologie inmiddels is en hoe snel deze zich op bijna onzichtbare manier verbetert. Anderzijds maakt elke rit met Autopilot zichtbaar hoe complex de voor ons schijnbaar eenvoudige handelingen als autorijden voor technologie in werkelijkheid zijn. En hoe “dom” en basaal de fouten van AI zijn. *Autopilot* brengt de ontwikkeling van AI dichtbij én laat meteen zien hoe ver weg de volledig zelfrijdende toekomst is. Leven in de tussentijd dus: onderweg naar geheel zelfrijdend, maar in de fundamentele onzekerheid of we dat niveau ooit zullen bereiken.

En dat heeft grote consequenties. Bijvoorbeeld in het denken over mobiliteit, die ik later nog nader zal uitwerken. Een toekomst met alleen maar volwaardige op niveau 5 zelfrijdende voertuigen is totaal anders dan een toekomst met deels gewone auto's, deels zo goed als zelfrijdende auto's, afgewisseld met een paar geheel zelfrijdende “auto's” (want het is de vraag of het woord auto daarvoor nog de lading dekt). In termen van ruimtelijk beslag, regels en de inrichting van het totale systeem van mobiliteit is dit niet een verschil in gradatie, maar een verschil tussen dag en nacht. Daarom is de vraag *hoe perfect* AI wordt geen detail, maar de kern van elk gesprek over AI. We zijn onderweg, maar de aard van de bestemming is fundamenteel onzeker. We weten *dat* we onderweg zijn, maar niet *waar naartoe* precies. Maar daar hebben we het vreemd genoeg heel weinig over. En dat is vreemd. Niemand slaat voor hij of zij met vakantie gaat de vraag over wat de bestemming van de reis eigenlijk is. Dat is immers van belang voor wat er in de koffer gaat, of er überhaupt een koffer mee gaat, wat een handig vervoermiddel is, et cetera. Het lijkt een stupide vraag, maar het is de meest basale vraag die we voor AI tot nu toe niet beantwoorden. En eigenlijk zelfs nauwelijks stellen. Dat is de vraag die in dit essay centraal staat. Omdat het volgens mij de enige écht belangrijke vraag is die rond AI te beantwoorden is. Als we in de tussentijd verkeren, dan is het minste wat we moeten weten (of wat we ons in ieder geval moeten afvragen) waar we naartoe onderweg zijn. En wat er onderweg nodig is om die bestemming veilig te bereiken? Naar die antwoorden zoek ik in dit essay.



## Leeswijzer

Zo gaat dit essay over de betekenis van AI voor het openbaar bestuur en de democratie, en over de manier waarop we onderweg naar de toekomst met AI kunnen omgaan. Allereerst bespreek ik wat voor mij de werkdefinitie van AI en digitale technologie is, en hoe AI en digitale technologie van toepassing zijn op de democratie en het openbaar bestuur. Vervolgens bespreek ik wat voor mij de essentie van democratie, bestuur en sturing vormt, en wat dat betekent in relatie tot AI. Hierna ga ik in op verschillende manieren van denken over toekomstige ontwikkelingen en op de manier waarop we de kansen op een gunstige uitkomst voor AI in het openbaar bestuur kunnen doen toenemen. Ik sluit af met een drietal strategieën voor het sturen van AI en digitale technologie, waarin het bestuur in elk van die strategieën een andere rol ten opzichte van technologie aanneemt.

## 2. Silicons en carbons: oude wijn in nieuwe zakken, of nieuwe wijn?

Technologie is een disruptieve kracht in het openbaar bestuur. Er zijn weinig mensen die dat zullen ontkennen. De praktijk van sturen en organiseren is in onze tijd ontegenzeggelijk anders dan twintig, dertig of vijftig jaar geleden. En ook de komende tien jaar zullen grote veranderingen laten zien. Wat zal precies de aard van de digitale disruptie zijn? Gaat het om een verandering van middelen in niet fundamenteel veranderde processen, of is de verandering meer fundamenteel van aard? Dat gaat dus niet om de vraag of veranderingen groot of klein zijn – ze zijn groot – maar om de vraag of technologie de fundamenten van het openbaar bestuur verandert. Spoiler alert: *ik denk (nog) van niet*. En ik zal uitleggen waarom.

Maar eerst is het handig om de cryptische titel van deze paragraaf te verklaren. Ik was bij een lezing van complexiteitwetenschapper Dave Snowden in Singapore, op een jaarlijks door de Singaporese overheid georganiseerd congres over toekomstverkenning. Overheidsbestuur gaat over het omgaan met complexiteit en ambigüiteit betoogde hij en daarvoor kun je twee soorten systemen inzetten: *silicons* en *carbons*. Carbons zijn wij; mensen. Opgebouwd uit koolstofverbindingen, die met hersenen en zintuigen complexiteit waarnemen, er patronen in menen te herkennen en op basis daarvan besluiten nemen en handelen. Daar tegenover staat de opkomende wereld van silicons, de microchips die zorgen voor steeds meer rekenkracht en die daarmee ook steeds beter in staat zijn om complexiteit te lezen en daarin te handelen. Silicons en carbons hanteren heel andere manieren van werken en ‘denken’, en hebben heel andere kwaliteiten. Silicons zijn heel goed in heel snel sequentiële reeksen data in binaire categorieën in te delen en daarmee ‘het beeld’ en ‘de opties’ te genereren. Silicons calculeren. Carbons kunnen niet goed sequentieel denken en al helemaal niet goed calculeren. Althans, het is een handeling die carbons heel veel tijd kost.

Carbons kunnen in plaats daarvan wel iets anders heel goed. Carbons kunnen als geen ander op basis van een brij van wat ze zien, weten, kennen, horen, denken, verwachten, overwegen, of vermoeden, *verhalen construeren* over wat er aan de hand is, wat er staat te gebeuren en wat daarvoor nodig is. Carbons hebben ook de wil, ambitie en noodzaak om te handelen. Ze zijn creatief en ondernemend. Ze nemen voor de realisatie van hun projecten en ideeën van alles samen, gooien het intuïtief of op basis van eerder opgedane ervaringen

op een samengestelde hoop, vullen alle onduidelijkheden daarbij naar eigen inzicht creatief in, en maken daar een eigen verhaal van. Zo komen ze in de meest onduidelijke, complexe, ambigue en voor hen vreemde situaties tóch razendsnel tot geïnformeerde en intuïtieve verwachtingen en keuzes. En die keuzes vormen de basis voor snel handelen, hoe complex en onoverzichtelijk de omstandigheden ook zijn.

Temidden van complexiteit blijken beide systemen goed te werken, maar wel ieder op hun eigen manier. Volgens de eigen route en met eigen kracht en tekortkomingen. De tussenstand daarbij is dat siliconen goed zijn in alles wat sequentieel calculeerbaar is, en carbons beter zijn in al het andere. En dan blijkt een rotonde of een over de weg wandelende voetganger voor de calculerende silicon-Autopilot van Tesla heel moeilijk en voor Henk-de-carbon-bestuurder een heel makkelijk en bijna achteloos te nemen object. Zelfs als Henk niet zit op te letten. Andersom is een voor de auto springende voetganger of plotselinge noodstop van een voetganger voor de silicon autopilot makkelijk en routinematig te ontwijken, en voor Henk-de-carbon-bestuurder heel moeilijk, zelfs als hij volledig geconcentreerd achter het stuur zit. De silicon trapt instant de rem in bij een obstakel, de carbon heeft tijd nodig om waar te nemen, te verwerken en actie te ondernemen.

Siliconen en carbons zijn manieren van *doen*, die niet alleen van toepassing zijn op autorijden, informatie zoeken of productiewerk in een fabriek. Het zijn ook verschillende manieren om tot overheidshandelen te komen. Centraal in overheidshandelen staat wat ik een vorm van een *besluit* noem. Elk overheidshandelen omvat “een besluit”. Of dat nu gaat om het slaan van een beschikking, het vaststellen van beleid, of het ondernemen van een handeling in de uitvoering, zoals het aanhouden van iemand op straat of het signaleren van een leerling die geen ontbijt lijkt te hebben gegeten (en dat al enkele weken lang). *Het besluit* is de kern van overheidshandelen. Het is als het ware het elementaire deeltje van het overheidsbestuur.

Bij elk besluit vindt een confrontatie plaats van een *feit* of fenomeen, al dan niet in context te beschouwen, en een (al dan niet formele) *norm*, waarop een beslisregel wordt toegepast. Deze vorm van afwegen vormt de basis van elk handelen door de overheid. Het is de basis van de bureaucratische organisatievorm. Daarbij zijn de toe te passen normen *politiek*, is de toepassing op het “feit” neutraal, en dient het feit zo geobjectiveerd mogelijk vast te stellen zijn. Het gaat dus (idealiter) niet om het loslaten van normen op gevoelens of vermoedens, maar op objectieve feitelijke fenomenen. Of iemand in aanmerking komt voor een uitkering is niet de uitkomst van de confrontatie van een gevoel over hoe iemand er financieel bij zit en een gevoel dat dat compensatie en tegemoetkoming verdient. In plaats daarvan zoeken we in

de bureaucratische organisatievorm naar een objectieve weergave van de financiële situatie, waar we de politieke norm – die eveneens geobjectiveerd is – op toepassen, waarna via de beslisregel volgt of er wel, geen of een beetje tegemoetkoming volgt.

Dit rudimentaire proces van het in de praktijk brengen van *besluiten* is de afgelopen decennia in de opeenvolgende golven van technologische vernieuwing ingrijpend veranderd. In eerste instantie door *bureaucratisering* van het bestuur. In de door Weber ontworpen bureaucratische organisatievorm werden de stappen van een besluit geheel ingevuld door *mensen*, door ambtenaren. Het proces was geobjectiveerd, in die zin dat de carbons in kwestie zoveel mogelijk op basis van feitelijke gegevens de normen toepasten op de ‘gevallen’ die zij voor zich kregen. Daarbij pasten ze eveneens geobjectiveerde normen toe op de concrete gevallen. Dat was letterlijk handwerk en ‘hoofdwerk’, uitgevoerd door mensen. De carbons zaten (bij afwezigheid van siliconen) stevig aan het roer, ingebed door procedures, regels, controles en standaarden.

De eerste technologische golf die op dit proces van invloed was, was *automatisering* en *informatisering*. Daarmee doel ik op de toepassing van technieken die het mogelijk maakten om gestandaardiseerde processen deels met behulp van ‘automaten’ uit te voeren en door middel van vroege computervormen steeds meer informatie op te slaan. Daarmee was de bureaucratie in staat om méér informatie te genereren en digitaal op te slaan en deze vervolgens te gebruiken voor het beschrijven en vastleggen van informatie over fenomenen, en daar geheel of gedeeltelijk geautomatiseerd beslisregels op toe te passen. Hierin gebeuren dus twee dingen. Allereerst het ingrijpend vergroten van de hoeveelheid beschikbare en behapbare digitale informatie over fenomenen en het (semi-)geautomatiseerd op die digitale informatie toepassen van beslisregels. De informatie wordt niet meer handmatig door mensen verzameld en vastgelegd, maar deels automatisch gegenereerd en vastgelegd in digitale informatiesystemen. Dezelfde informatie, maar meer. En dezelfde typen beslisregels, maar geautomatiseerd en daardoor meer en sneller.

Informatisering en automatisering via *digitalisering* zorgden voor grote veranderingen in de werking van de overheidsbureaucratie. Feitelijk ging dat om het digitaliseren van dezelfde bureaucratische processen die eerst ‘mensmatig’ door carbons werden uitgevoerd. Informatie over een fenomeen wordt door carbons in het digitale systeem ingevoerd en gecombineerd met reeds bestaande digitale informatie aldaar. Waarna het systeem het via de digitale informatie geconstateerde fenomeen confronteert met de digitaal geprogrammeerde norm. Via een beslisregel komt dan een *besluit* tot stand.

Informatisering en automatisering maken dat die besluiten via de route van de siliconen op veel grotere schaal kunnen worden genomen. Vaak gaat het ook sneller. En informatisering en automatisering betekenen dat langzaam maar zeker een begin kon worden gemaakt met het verzamelen van systematische data over maatschappelijke fenomenen en individuen. Daarmee veranderde al dan niet bedoeld ook de aard van de data. Informatisering betekent digitalisering, en dat vereist dat maatschappelijke fenomenen in binaire data worden omgezet. Het is 0 of 1. Daarmee werd de verzamelde en opgeslagen informatie dus ook steeds minder ambigu, omdat de systematiek geen ruimte laat voor accenten, nuance, of onzekerheid die voorbij de 0 of 1 gaat. Digitalisering is een binair systeem en vereist altijd voor alles een keuze tussen 0 of 1. Hoezeer we ook oprecht worstelen met pogingen om meer ambiguïteit in de datavisualisatie en presentatie te brengen. Uiteindelijk zijn toch de binaire categorieën nodig, anders kan het systeem het materiaal simpelweg niet kwijt. Wel maakt de met de tijd steeds verder groeiende opslagcapaciteit en processorsnelheid het mogelijk om complexe maatschappelijke fenomenen in steeds *méér* binaire grootheden te vatten: in steeds meer 0-en en/of 1-en. Zo kan toch enige binaire nuancering plaatsvinden. Het binaire systeem kan gebruikt worden om eindeloos veel grijsstinten te benoemen, in plaats van alleen *óf* zwart (0) *óf* wit (1). Zo kan het systeem veel *méér* binaire informatie dan eerst opslaan en ontstaat zo ook meer ruimte voor nuance en kleurverschil. En omdat parallel met de groei in opslagcapaciteit ook het vermogen om te rekenen exponentieel toeneemt kan het systeem ook nog steeds in een oogwenk die informatie lezen en bewerken. Het silicon-proces is daarbij nog steeds sequentieel, maar kan die sequentiële stappen *zó* snel en grootschalig uitvoeren, dat het fysiek beperkte carbon-oog het als zo goed als 'instant' ziet.

Zo werd de eerste stap van de ontwikkeling van carbons naar siliconen in het openbaar bestuur gezet. Van menselijke (carbon) vergaring, opslag en invoer van data en menselijke toepassing van norm en beslisregels, naar grootschalige digitale (silicon) opslag en vergaring van digitale data, met geautomatiseerde digitale toepassing van normen en beslisregels. De carbon staat hier nog steeds aan het roer, want de mens programmeert ten behoeve van de bureaucratie een grote lopende band, die routinematig grote hoeveelheden informatie confronteert met geprogrammeerde normen, die daar via eveneens geprogrammeerde beslisregels conclusies aan verbindt. En die lopende band zet ook vaak meteen automatisch de digitale vervolgstappen, zoals het uitprinten van een brief en het opslaan van de uitkomst in het informatiesysteem. Simpel gezegd, wat eerst de carbon zelf deed doen nu de siliconen van het geautomatiseerde gedigitaliseerde systeem. Daarmee werd de bureaucratie in zekere zin gesublimeerd, omdat in tegenstelling tot de *mensen* (ambtenaren) de informatiesystemen en automaten niet van regels

afwijken en geen eigen interpretaties geven. Ze kunnen geen opvolging geven aan een ‘niet pluis gevoel’ of vage vermoedens over zwakke niet gerubriceerde signalen. Automatisering en informatisering leidden daarmee dus tot standaardisering en disciplineren. De ruimte rondom de regels en objectieve informatie nam er sterk door af. De inzet van siliconen past dus heel goed in het bureaucratische ideaaltypen van Weber. Hoewel Weber bij het ontwerp van zijn bureaucratietheorie nog geen siliconen kon voorzien, is het bijna een vervolmaking van de door hem bedoelde bureaucratische organisatievorm en de daar onderliggende principes.

Daarnaast had digitalisering wel nog een belangrijk tweede gevolg, dat wellicht minder was voorzien en al helemaal niet was bedoeld. De voor het systeem benodigde binaire informatie is *performatief*. Dat betekent dat het systeem niet alleen informatie binair opslaat, maar ook dwingt om de maatschappelijke omgeving binair te bezien en te benaderen. In eerste instantie wordt rijke waarneming ‘omgezet’ naar binaire code, waarin het nodige verloren gaat. Maar vervolgens gaat het systeem vanuit binaire code ook de wereld benaderen en actief op zoek naar gemakkelijk in binaire code om te zetten informatie, om de vaak meer complexe maatschappelijke fenomenen te representeren. De binaire code neemt zodoende de representatie van het fenomeen over en reduceert rijke en inherent ambiguë of dynamische fenomenen tot binaire categorieën. Tot uiteindelijk de fenomenen zichzelf ook ‘meer binair gaan gedragen’. Daarmee bedoel ik dat ze bewust hun meer complexe identiteit in binaire vorm naar buiten brengen, op zo’n manier dat de kans op een positieve confrontatie met de norm er groter door wordt. Zo heeft informatisering veel verdergaande gevolgen dan het alleen maar ‘opslaan’ van informatie op een bepaalde manier. Het is ook van grote invloed op de manier waarop het systeem maatschappelijke fenomenen benadert en hoe vervolgens die maatschappelijke fenomenen ook zichzelf gaan gedragen en extern gaan representeren. Binaire systemen maken als het ware de wereld méér binair, hoe oneigenlijk dat voor veel maatschappelijke fenomenen – zoals een wijk, een cultuur, een gezinsleven, een menselijke gedachte, de aard van een ontmoeting, een liefdevolle maar spanningsvolle relatie – ook is.

De gevolgen van digitalisering en automatisering zijn dus groot, maar de fundamentele opbouw van het proces is er opvallend genoeg niet wezenlijk door veranderd. Wel zijn de verhoudingen tussen siliconen en carbons steeds meer verschoven, waarbij carbons nog ‘aan de knoppen zitten’ maar in de praktijk de miljoenen dagelijkse micro-besluiten van siliconen de maatschappelijke praktijk vormgeven. Een volgende stap in deze ontwikkeling en de groei van de rol van de siliconen is de toevoeging van algoritmen aan de mix. Of beter gezegd, meer uitgebreide algoritmen. Wat we hier beslisregel

genoemd hebben - ‘als *dit*, dan *dat*’<sup>1</sup> - is op zichzelf een vorm van een algoritme. De beperking in computercapaciteit maakte lange tijd dat alleen eenvoudige algoritmen goed geprogrammeerd en toegepast konden worden op de beschikbare data. De andere beperking was dat vanwege diezelfde technische begrenzingen de hoeveelheid data nog steeds te beperkt was om meer ingewikkelde, uitgebreide algoritmen op los te laten. Door de technische innovaties in opslag- (volume) en processorcapaciteit (snelheid) werd het mogelijk om steeds meer data te genereren en op te slaan, deze sneller te doorzoeken en in dezelfde tijd met meer veelzijdige en uitgebreidere beslisseregels te confronteren. Zo werd het bijvoorbeeld mogelijk om verschillende datasets aan elkaar te koppelen en daarin verbanden naar voren te brengen. De ontwikkeling van meer uitgebreide algoritmen maakte bovendien mogelijk dat algoritmen onderweg zelf “vervolgstappen” in de confrontatie met data konden gaan zetten. Bijvoorbeeld door gericht door te zoeken op bepaalde delen van de data en zelf geprogrammeerde afslagen in de zoektocht te nemen. Daarmee werd de techniek om de data te ‘lezen’ en te doorzoeken rijker en nam ook de mogelijkheid tot koppeling van bestaande databronnen aan elkaar verder toe.

Deze ontwikkeling liep ongeveer synchroon met de exponentiële groei van de data zelf. De belangrijkste oorzaken hiervan liggen in de verdergaande digitalisering van de samenleving. Die maakt dat steeds meer processen via data verlopen en er simpelweg op veel méér plekken, door veel méér silicons, veel méér data wordt geproduceerd. Het gaat dan niet meer om het omzetten van fysieke processen naar representatie in de vorm van geproduceerde of ingevoerde data: de processen zijn zelf data geworden. Net zo belangrijk is de enorme ontwikkeling in *silicone sensorcapaciteit*. Data wordt niet zozeer meer door carbons waargenomen, omgezet en ingevoerd in silicon-systemen, maar silicon-apparaten registreren zelf, direct, in de vorm van data wat er buiten gebeurt. Die data is voor de carbon-waarnemer ook niet goed navolgbaar en leesbaar meer. Probeer maar eens met het eigen blote oog een database ‘te lezen’. Een klassieke kwikthermometer meet fysiek wat de temperatuur is en geeft deze in carbon-termen aan op een af te lezen schaal. De persoon – een carbon – kan die waarneming als datapunt vervolgens invoeren in het systeem. Digitale sensoren doen meteen al zelf een digitale registratie: wat we in carbon-begrippen een waarneming zouden noemen. Die waarneming *is* data, die niet wordt omgezet *in*, maar zelf de vorm van

---

1 Let op: het ‘probleem’ van algoritmen en digitalisering en automatisering zit dus ook niet in de ‘dan’, maar in de ‘als dit’ en de ‘dat’. Het algoritme vereist een binaire ‘als’ die dan via de beslisseregels kan worden omgezet in een al even binaire ‘dat’. Het verlies aan menselijke maat is daar aan te wijzen: het onvermogen om rijk, ambigu en in essentie contextueel te kijken naar de ‘als’, naar wat er speelt, om daar een passende handelingsoptie aan te verbinden.

data heeft. Op een centraal netwerk aangesloten thermometers generen zo automatisch data over het klimaatsysteem over de wereld. Die data wordt niet meer afgelezen en ingevoerd, maar is in zichzelf al gedigitaliseerd en voor carbons alleen nog via een daarvoor geprogrammeerde interface af te lezen. Niet de carbons voeren informatie voor de siliconen in, de carbons “vragen” via programmatuur de siliconen om de informatie waarover zij beschikken ook voor carbons leesbaar te maken. Zonder die programmatuur is de informatie voor de carbons niet meer te lezen of te begrijpen.

Met de exponentiële afname van de kosten en de eveneens exponentiële capaciteit van silicone sensoren, kwamen er steeds meer digitale sensoren die van alles en nog wat in het leven omzetten in data. Niet bij wijze van specifieke aanvullende handeling, maar in de kern van wat *sensing* is. De *sensing* is gedataficeerd en heeft grotendeels de waarneming door carbons vervangen. Daarbij is veel handelen in zichzelf al data geworden. We zetten geen stap, doen geen click, slaan geen hartslag en halen geen adem zonder dat het als datapunt door devices en sensoren als data wordt geregistreerd. Vroeger noteerde mijn moeder bij elke tankbeurt in een klein schriftje het aantal gereden kilometers en de hoeveelheid liters getankte benzine. Zo creëerde zij data over het benzineverbruik van de auto en misschien onbedoeld ook haar eigen rijgedrag. Tegenwoordig *is* de rit een permanente data-administratie en lezen hulpdiensten of opsporingsdiensten na een ongeval simpelweg de auto uit om te weten wat er is gebeurd. Het rijden *is* de data. En zo is dat met heel veel grote en kleine activiteiten in het dagelijks leven van individuen, maar ook van hele gemeenschappen en groepen, gebeurd. Het Smart City concept is daarop gebaseerd: het via ongelimiteerde *sensing* vastleggen van alle activiteit in een ecosysteem (een buurt, wijk of stad) en die data gebruiken voor een gedataficeerd inzicht in wat die gemeenschap beweegt. Met als volgende stap, voorbij de wearables, ook *sensing* in het lichaam, die de siliconen letterlijk *in* de carbons brengt.

De data die deze sensoren genereren is soms besloten en in private handen, soms open, soms van de overheid zelf, en vaak is simpelweg niet duidelijk wie de eigenaar is.<sup>2</sup> Duidelijk is wel dat de verzamelde hoeveelheid, en de mogelijkheid om via internet data onderling te verbinden en uit te wisselen, de pool van beschikbare data oneindig groot heeft gemaakt. Dat proces gaat door en groeit exponentieel. Zo is wat we *big data* noemen ontstaan. Gigantische bestanden bron-data, nog niet van betekenis of orde voorzien, die door middel van zoeken, ordenen en verbanden leggen kan worden

---

2 En dit is onderwerp van juridische strijd (van wie zijn bijvoorbeeld de data die een Tesla of andere auto genereert) en verhit politiek debat (zoals bijvoorbeeld rond het elektronisch patiëntendossier en andere medische gegevens).



geïnterpreteerd in al dan niet zinvolle betekenissen. Bijvoorbeeld over kijkgedrag van consumenten, de ontwikkeling van tumoren, of de patronen in de trek van vogelsoorten. Daarbij is het belangrijk om te realiseren dat wat we nu onder *big data* verstaan we over decennia glimlachend zullen benoemen als een prehistorische poging daartoe. Zoals we nu op de iPhone melancholisch filmpjes terugkijken van de Atari spelcomputer of de Commodore64. Toen hightech, nu oldschool. Met als Droste-effect dat de iPhone en de Big Data van nu, de Commodore64's van de nabije toekomst zullen blijken. En de Tesla Model S ooit, en waarschijnlijk sneller dan we denken, als een oldtimer of *classic* zal aanvoelen. Mensen hebben de behoefte om de huidige situatie als eindpunt of hoogtepunt in de geschiedenis te ervaren. Als de culminatie en sublimering van alles wat er daarvoor is gebeurd. Maar de realiteit is minder romantisch: de state of the art en high-tech van nu is altijd het 'hopeloos achterhaald' en de romantiek van de toekomst. Met daarbij ook de voor dit essay cruciale aantekening, dat veel toekomstmuziek uiteindelijk niet uitkomt en ook exponentiële ontwikkeling uiteindelijk toch vaak afvlakt: denk aan gratis energie uit kernfusie, de definitieve doorbraak van waterstof, maar ook het daadwerkelijk zelfrijden van auto's. Het zijn technologieën die al jaren worden voorspeld, maar waarvan de allerlaatste stap toch lastig blijkt te zijn.

Uitgebreide algoritmen, grote databestanden en snelle reken capaciteit hebben elkaar nodig. Letterlijk. Ze kunnen zonder elkaar niet bestaan. De databestanden zijn te groot om met simpele algoritmen zinvol te doorzoeken, en die combinatie vereist dus enorme reken capaciteit. Die uitspraak veronderstelt wel dat het algoritme met een vooraf bepaalde strategie en beoogde doelstelling is gebouwd. Dat een algoritme is gemaakt om *iets* te vinden in een hele grote hoeveelheid data. Alsof het algoritme de speld in de hooiberg zoekt. Met het juiste algoritme en voldoende snelle reken capaciteit is de vooraf gedefinieerde speld uiteindelijk door de silicons in de data-hooiberg te vinden: hoe groot deze ook is. In deze vorm gebruiken carbons de silicons om voor hen te vinden wat ze zoeken: de speld in de hooiberg gaat uit van mensen die weten dat ze een speld nodig hebben, hoe deze er uit ziet én dat deze zich ergens in de hooiberg bevindt. En wat überhaupt een hooiberg is. Maar, dat stelt de mens centraal bij algoritme-ontwikkeling, bijvoorbeeld het vermogen van mensen om patronen en verbanden tussen bepaalde delen data of datasets vooraf te voorzien. En om daaruit eigenschappen van *speld* en *hooiberg* te benoemen. De menselijke creativiteit heeft daarbij een centrale, maar daarmee inherent ook beperkende factor. Algoritme-ontwikkeling werkt echter net zo goed andersom; algoritmen kunnen en worden zo gebouwd, dat ze in staat zijn om in de hoeveelheid data *zelf* verbanden te vinden. Met andere woorden, ze hebben een hooiberg nodig om te ontdekken wat de speld is die zich erin bevindt – en wat die speld precies is. Het vooraf bedachte relevante verband wordt niet in de data gevonden, in de data wordt

gevonden wat eigenlijk het relevante verband *is* en zou *moeten zijn*. En welke nóg interessantere spelden door middel van koppelingen met andere grote databestanden gevonden kunnen worden. Dat alles kan toevallig in lijn zijn met wat de carbon toch al zocht, maar hier kunnen de siliconen ook een heel ander onverwacht en ongedacht verband vinden. Er is veel discussie over de vraag of siliconen daarbij in staat kunnen zijn om tot betekenis (*sense of meaning*) te komen, of dat ze niet meer kunnen dan reproduceren wat ze in interactie van carbons hebben herkend als “betekenisvol”. Met andere woorden, kunnen siliconen zelf nieuwe categorieën van interesse benoemen, of beperken ze zich tot het concept “interessant” dat ze carbons hebben ‘zien’ gebruiken?

In beide gevallen kunnen siliconen zelf op zoek gaan naar verbanden die carbons er niet zelf vooraf in geprogrammeerd hebben. Dat kan met *gefixeerd zoekende algoritmen* gebeuren, maar algoritmen kunnen ook zo worden ontworpen dat ze zich onderweg aanpassen aan de data die ze vinden en daarmee andere zoekrichtingen inslaan. *Machine learning* gaat over algoritmen die zichzelf trainen en wijzer maken en die ook zelf nieuwe en originele afslagen kunnen nemen. Ze doen dat (nog) niet door eigen creativiteit, maar door generatief de bij carbons “afgekeken” patronen en concepten verder uit te werken, te herhalen en te verdiepen. Door dit te doen veranderen de algoritmen in de interactie ‘zichzelf’, waardoor ze al zoekend op basis van wat ze vinden ook veranderen wat ze zoeken en zullen vinden. De siliconen vinden dan dus niet alleen spelden en hooibergen, ze veranderen en vinden daarbij ook zichzelf. Ze worden niet wat de carbons van ze willen, maar wat ze van zichzelf maken. En vervolgens zoeken en vinden ze wat ze “zelf” interessant of anderszins van belang en de moeite waard achten.

Big data is op deze manier de grote aanjager achter de belangrijkste verdienmodellen van onze tijd. Of beter gezegd, venture capital heeft big data en technologie ontdekt als verdienmodel en investeert daar grif in. Technologiebedrijven, maar ook platformbedrijven zoals Uber, en social media bedrijven, bieden diensten gratis of ‘goedkoop’ aan, in ruil voor gebruikersdata. Zo bouwen ze via hun diensten miljarden datapunten die ze vervolgens kunnen gebruiken of verkopen. Netwerkstructuren hebben zodoende een dubbele werking. Ze *horizontaliseren* onze wereld, in die zin dat ze iedereen die wil in contact brengen met eindeloze hoeveelheden informatie, anderen, muziek, video en andere content. Dat democratiseert en maakt van elke gebruiker en consument in potentie een aanbieder en een producent. Ze zorgen zodoende voor toegang van iedereen tot alles – en tot iedereen. Daarmee maken ze verbinding en contact kosteloos en doorbreken ze de verticale hiërarchieën en machtsposities die dat contact eerder reguleerden, zoals gatekeepers, tussenpersonen, redacties en betaalde aanbieders.

Tegelijkertijd doen ze dat alles niet belangeloos, maar als onderdeel van een eigen verdienmodel en *monopolievorming*. Platformbedrijven bouwen via de interacties op hun kanalen extreme dataposities op, die ze commercialiseren én die ze nodig hebben om hun platform verder te verbeteren en te versterken. Dat zorgt voor *verticalisering* en *monopolisering*, in die zin dat steeds verdere schaalvergroting plaatsvindt én dat bedrijven proberen steeds meer gebruikers in hun platforms binnen te sluiten. Zo waaiert de markt niet breed uit, maar groeit deze aangejaagd door de noodzaak van stapeling van data steeds meer uit naar een oligopolie van enkele dominante partijen. Die partijen en hun platforms zijn te groot om elkaar uit te schakelen en tot een algeheel monopolie te komen, anders hadden ze het zeker gedaan. Vreemd genoeg voelt voor dagelijks gebruikers deze oligopolisering heel anders: als vrijheid en onafhankelijkheid. En voor sommige anderen voelt het weliswaar als gedeeltelijke gevangenschap, als een dringend beslag op hun vrijheid, maar zij wegen nog altijd de voordelen positiever dan de nadelen. Zo leiden platforms op de korte termijn, dichtbij gebruikers en in dagelijkse handelingen, tot grote democratisering en horizontalisering. Op de langere termijn en op het meta-niveau betekenen ze een enorme verticalisering en oligopolisering. De ruimte en vrijheid die ze bieden in de directe levenssfeer gaat gepaard met permanente *capture* op meta-niveau. De carbons zijn hier de productiemiddelen van de siliconen geworden, die op hun beurt als kapitaal in handen zijn van een kleine groep carbons.

Zo is er sprake van *silikonialisering*. Carbons, gebruikers en consumenten, merken daar in de praktijk niets van. Sterker nog, in tegenstelling tot eerder fysiek gekoloniseerde volkeren ondervinden ze op korte termijn veel positieve gevolgen. Muziek en video waren in de geschiedenis nog nooit zo goedkoop, toegang voor producties was nog nooit zo gelijk, het speelveld lag nog nooit zo open. Contact was nog nooit zo makkelijk te leggen en te onderhouden. De transactiekosten van veel zaken in ons dagelijks leven, zoals het zoeken naar informatie, zijn vaak nihil. Maar de prijs van die lage transactiekosten is wel dat onder de radar, of erboven, al die transacties grondstoffen zijn voor dataficering door een klein aantal partijen – wat de mensen die bij die transacties betrokken zijn tot productiemiddel van dataficering en van die bedrijven maakt. *If it's free, you're the product*, luidt de oude internetwijsheid. Inmiddels is die waarheid verder ontwikkeld en zijn gebruikers niet alleen het product, maar ook de grondstof en productiefactor geworden (zelfs nu veel van die producten allang niet meer gratis zijn). Gebruikers zijn tegelijkertijd de blijde afnemers van goedkope energie én de steenkool die in de kolencentrale wordt verbrand – en als ze pech hebben worden ze ook nog eens ziek van de uitstoot van dat verbrandingsproces. Het is een in de geschiedenis uitzonderlijke symbiose van siliconen en carbon, die tegelijkertijd elkaar én gebruiken én gebruikt worden. Uitbuiting was, om het cru te zeggen,

voor mensen nog nooit zó leuk en aangenaam. Silicons en carbons hebben hier een relatie zoals in het trappenhuis van Escher, die inherent in elkaar overloopt, zonder begin en zonder einde.

Recent is deze beweging van dataficering in een volgende exponentiële curve terecht gekomen. De ontwikkeling van *large language models* (LLM's), met *ChatGPT* als de meest bekende toepassing ervan, zijn inmiddels mainstream geworden en beginnen impact te ontwikkelen op het dagelijks leven. Eerder bestonden ze al in expertdomeinen, zoals het juridisch en het medische domein. Met language software was het bijvoorbeeld mogelijk om alle internationale wetgeving bij te houden en te interpreteren, bijvoorbeeld om internationale handelsbedrijven goed te kunnen adviseren over de regelgeving waaraan ze in verschillende landen moeten voldoen en de standaarden die daar gelden. Voor carbons is het onmogelijk om alle nieuwe wetten, regels en jurisprudentie bij te houden. Voor silicons is dat relatief eenvoudig. Als silicons eigenstandig leren lezen, en er geen (re)codering meer nodig is, dan ligt de wereld letterlijk open voor upload. In tegenstelling tot carbons zijn voor silicons het kunnen lezen en spreken/schrijven van een taal ongeveer hetzelfde en gemakkelijk te combineren: kunnen lezen *is* voor silicons kunnen spreken. Die stap was dus simpel gezet, ChatGPT schrijft moeiteloos emails, instructies, werkstukken, spreekbeurten en het programma voor de volgende heidag van het bedrijf.

Dat spreken en zelf schrijven gaat inmiddels ook al verder dan alleen het simpelweg regenereren van wat eerst door iemand anders al eens letterlijk zo is gezegd en/of is gepubliceerd. Op Wikipedia is (van) alles te vinden, maar dat is daar door carbons op gepubliceerd en wordt in een netwerk van gebruikers permanent bijgewerkt, onderhouden en geredigeerd. Het is online digitaal beschikbaar, maar het is er door carbons geplaatst: het is een groot door carbons gevuld en gegeneerd informatiesysteem. Bij LLM's werkt dat anders. Die zijn in staat om de patronen en systemen van de taal te interpreteren: ze weten dus niet alleen welke informatie er wordt gedeeld, maar ook wat er wordt gezegd en wat daarmee wordt bedoeld. En toen ze dat eenmaal "begrepen" was de volgende stap eenvoudig te zetten, namelijk het zelf verzinnen van samenhangende taal en tekst. LLM's delen niet alleen informatie die door carbons eerder is opgetekend – zoals een zoekmachine doet –, ze creëren zelf eigen talige content. Ze maken zelf hun eigen verhaal en kunnen daarbij net als carbons verschillende feiten en gegevens bij elkaar brengen in de patronen van taal en verhaal. Alleen dan oneindig veel sneller. En over elk mogelijk onderwerp.

Dat is nog steeds maar het begin en de volgende stap is alweer onderweg naar brede maatschappelijke adoptie. LLM's interpreteren inmiddels al voorbij de

*tekst* en zijn in staat om allerlei andere vormen voor betekenisoverdracht te herkennen: beeld, geluid, spraak, beweging, kunst. Silicons *ver-talen* en *ver-teksten* wat ze waarnemen, net zoals carbons doen, in verhaallijnen waarmee ze heel verschillende vormen met elkaar kunnen verbinden. Dat is niet alleen maar een nieuwe afslag voor LLM's, het betekent dat voortaan niet alleen *teksten* onderling uitwisselbaar zijn via LLM's, maar dat *alle* informatie onderling uitwisselbaar is. Dat is een exponentiële verandering. Alle informatiebronnen, alle uitdrukkingsvormen, alle manieren van het overbrengen of vastleggen van betekenis kunnen via LLM's min of meer direct met elkaar worden verbonden. En daarmee is in één klap een onvoorstelbare pool van reeds bestaande informatie met elkaar verbonden én kan in diezelfde klap meteen ook in de toekomst alle informatie aan elkaar worden gekoppeld. Het betekent dat innovatie en verbetering van de LLM's niet meer per sector hoeft te gebeuren, maar dat er één gedeelde taal is ontstaan om alle domeinen te interpreteren, te begrijpen en nader te bewerken. Zo is als het ware een *Rosetta Stone* voor alle informatie beschikbaar gekomen: voor *alle* mogelijke informatie. Daarmee is de pool data en informatie op slag eindeloos geworden. De gevolgen die we op 'kleinere' schaal al hebben gezien met big data zullen we de komende tijd als gevolg van LLM's gaan zien, ook al is de praktijk nu soms nog net zo onbeholpen als de Tesla op de rotonde. LLM's gaan onze samenleving ingrijpend veranderen. We kennen ChatGPT van de handige antwoorden die het geeft op onze vragen: maar de volgende stap is dat 'Chat' zelf de vragen stelt en/of vertelt wat we eigenlijk zouden moeten weten en vragen, en wat misschien ook wel niet.

Wat betekenen deze ontwikkelingen voor het overheidsbestuur? De combinatie van algoritmen en dataficering zorgen voor drie belangrijke veranderingen in het proces van sturing. Allereerst neemt de pool van data over fenomenen exponentieel toe. Er is veel en veel meer informatie beschikbaar om feiten over fenomenen uit te destilleren en daarop vervolgens de norm toe te passen. Er is voor de overheid veel meer te weten en te kennen over maatschappelijke fenomenen en daar zijn de normen op toe te passen. Sturing en interventies kunnen daarmee méér, groter en in potentie preciezer worden toegepast. Er is simpelweg méér om mee rekening te houden en de exponentieel gegroeide snellere hardware maakt dat ook fysiek mogelijk.

Daarnaast geldt ten tweede dat de informatiesystemen zelf verbanden 'ontdekken' die anders niet *gedacht* waren. Dat is een wezenlijke verandering. Voorheen ging men actief en strategisch op zoek naar *vooraf bedoelde verbanden* of gevolgen, nu kunnen systemen *zelf* verbanden vinden, die niet gezocht, bedoeld of zelfs gedacht waren. Waarbij soms ook niet goed reproduceerbaar is welke wegen het algoritme onderweg heeft genomen om op de verbanden uit te komen, althans niet als dat niet expliciet is

geprogrammeerd. Zo is het datasysteem voor het eerst in staat om op uitkomsten te komen die *niet* vooraf door de mens zelf zijn ingestopt en ook voorbij de in de bouw/het ontwerp van het systeem aangebrachte begrenzungen te gaan. Zo kan het systeem bijvoorbeeld patronen ontdekken in de manier waarop mensen door de stad bewegen, waarom ze bepaalde keuzes maken, waar verdwenen belastingplichtig vermogen zich bevindt, of waar de overheid kansen laat liggen bij de aanpak van een maatschappelijk probleem. Siliconen zien meer en sneller en daardoor zien ze wellicht ook andere zaken en patronen. Die de waarneming of het voorstellingsvermogen van door het verleden en hun eigen padafhankelijkheden en bias getekende carbons, met hun beperkte cognitieve vermogens en onvolledige informatie, simpelweg niet *kunnen* of willen zien.

De derde consequentie van de opkomst van de siliconen heeft betrekking op het politieke in het overheidsbestuur. Siliconen kunnen bewegingen, patronen, kwesties, problemen en andere zaken waarnemen die de carbons die ze hebben gebouwd simpelweg niet zien. Dat kan toevalligerwijs vallen binnen eerder door carbons afgesproken politieke kaders, normen en regels. Maar waarschijnlijker is dat siliconen nieuwe politieke kwesties aanboren en markeren als “belangrijk” of “prioritair”. Hoe beter siliconen zoeken, hoe vaker ze zullen stuiten op zaken die politiek zijn en die *politiek maken* vereisen. De carbons gebruiken de technologie dan niet alleen meer voor de vervulling van de eigen politieke agenda, de siliconen *agenderen* zelf ook. Zij maken mede de politiek. Dat geldt voor problemen, maar ook voor potentiële oplossingen die ze genereren. Als systemen kunnen zoeken en vinden voorbij de vooraf geprogrammeerde grenzen, dan is de kans groot dat de oplossingen waarmee ze komen voorbij de bestaande politieke orde en begrenzing liggen. Wat nou als AI erin slaagt om de door sommigen in het sociale domein gedroomde “vroegsignalering” te vervolmaken? En wat als AI in de ruimtelijke planning “ontdekt” dat door bepaalde toegewezen vergunningen er allerlei kansen worden gemist: is dat dan voldoende aanleiding voor gedwongen onteigening? En wat als AI het perfecte spreidingsmodel voor windmolens en AZC’s heeft gevonden, inclusief de meest geschikte locaties – moeten daarvoor de lokale politieke besluiten en broze maatschappelijke consensus wel of niet worden opgebroken? Wat is, met andere woorden, de politieke status én statuut van wat de siliconen ongevraagd voor ons vinden en ontdekken? AI vereist niet alleen politieke begrenzing, maar zal ook vragen om politieke richting en nieuwe politieke standpuntbepaling, zoals overigens heel gewoon is bij de introductie van nieuwe technologie. Basaal genomen gaat dat over het primaat: zijn en blijven carbons boven de siliconen staan, of krijgen of nemen siliconen eigen politieke ruimte?

In het verre verlengde van deze drie consequenties ligt nog een heel belangrijke mogelijke vierde: *de silikolonisering van het politieke zelf*. Tot nu toe is in al die majeure veranderingen in het overheidsbestuur één ding hetzelfde gebleven: het antwoord op de vraag *wie stelt de norm?* Die is en blijft geformuleerd door de politiek, door mensen, door carbons. De norm is in decennia van oprukkende automatisering, digitalisering en dataficering het laatste bastion waar de carbon regeert. De data waarmee de norm wordt geconfronteerd is oneindig veel groter, en de beslisregel kan veel meer stappen en slagen omvatten dan eerst. Maar de norm van wat goed is en wat niet, wat wenselijk is en wat niet, en wat bedoeld is en wat niet, verandert daardoor niet. Tegelijkertijd maakt de technologie de waterscheiding tussen feitelijke informatie en politieke normstelling steeds lastiger houdbaar. Ten eerste, omdat het natuurlijk twijfelachtig is dat in geen van al die zoekstappen géén normatieve elementen zitten. Impliciet is wat siliconen doen op allerlei manieren al normatief en daarmee politiek. Die normativiteit zit al in de algoritmen zelf, maar zeker ook in de performatieve gevolgen ervan. En ten tweede, omdat nieuwe koppelingen van data en zoekstrategieën leiden tot nieuwe thema's. Ze genereren nieuwe spelden uit de zelf gevormde hooiberg, waarvoor nog geen normen zijn afgesproken. Daarmee drijft technologie ook tot nieuwe politieke uitspraken en stuurt het politieke processen in plaats van dat het deze volgt. Zo is het proces van silikolonisering van de politiek wellicht toch al meer gaande dan eerst. Maar dan in het gunstige geval impliciet en anders expliciet en bewust verborgen. En dan hebben we het nog niet eens over de politieke invloed in democratische processen die we recent in westerse democratieën hebben gezien. Digitalisering maakt extreem effectieve en vaak twijfelachtige beïnvloeding van kiezers mogelijk, die samen met desinformatie en nep-nieuws een reële bedreiging voor de democratische politiek en samenleving vormen. Maar daarover hierna meer.

In ieder geval op papier, is het onderscheid tussen de politieke norm en de geautomatiseerde, gedigitaliseerde en gedataficeerde vorming van overheidshandelen ('besluiten') nog steeds van kracht. Pogingen om algoritmen te registreren en te controleren passen ook in die opvatting, bijvoorbeeld in de vorm van de Algemene Rekenkamer die algoritme-control wil uitvoeren en uitvoeringsorganisaties die worden gecorrigeerd op de manier waarop zie via algoritmen en data tot "risicobeoordelingen" komen. Maar het is de vraag of dat onderscheid in de praktijk houdbaar is. Vanwege de enorme ontwikkeling en kracht van elk van de drie genoemde oorzaken – 1. normativiteit in de algoritmen en data, 2. performativiteit, 3. nieuwe kwesties – lijkt het erop dat er weliswaar politieke sturing door carbons *plaatsvindt*, maar dat digitalisering die inmiddels tegelijkertijd overstijgt en dat siliconen steeds meer eigen politieke ruimte innemen. In de nu volgende delen zal ik nader onderzoeken

of die positie houdbaar is en hoe deze, indien gewenst, de komende periode houdbaar kan blijven.

Om die houdbaarheid nader te onderzoeken verdiep ik de rol, invloed en werking van AI langs twee lijnen. Ik kijk eerst hoe AI zich verhoudt tot de specifieke eigenschappen en kenmerken van de *democratische rechtsstaat*. Hier spiegel ik de praktijk en het vooruitzicht van AI dus aan de normatieve kaders die de democratische rechtsstaat biedt. Vervolgens kijk ik hoe AI zich verhoudt tot de meer praktische en bestuurskundige vraag hoe *sturing en bestuur in de praktijk* vorm krijgen en hoe AI kan worden ingezet om in de praktijk publieke waarde te realiseren. En welke beperkingen zich dan voordoen. Dat gaat dus over de vraag hoe AI zich in de bestuurspraktijk manifesteert en wat daarvan denkbare gevolgen zijn.



# 3. AI in de democratische rechtsstaat

## Politiek, rechtsstaat en AI

Politiek en *democratie* gaan over gezaghebbende toedeling van waarden. Dat wil zeggen, over het vinden van een gezaghebbend antwoord op de politieke vraag *wie wat krijgt* (en waarom)? Politiek gaat dus over verdeling. Dat kan materiële verdeling zijn, bijvoorbeeld over de verdeling van middelen of kapitaal. Het kan ook gaan over immateriële of normatieve verdeling, over wie er gelijk krijgt, wie de ruimte heeft en wie zijn voorkeuren meer of minder bekrachtigd ziet. De keuze voor een progressief belastingstelsel boven een 'vlaktaks' is een materiële verdeling: de ene groep betaalt meer, de ander minder. De keuze voor beperking van het recht op abortus of de keuze voor het al dan niet afbreken van een protestmars is een normatieve verdeling: het één krijgt ruimte, het ander niet. In een democratie worden keuzes over verdeling bepaald door het realiseren van parlementaire meerderheden – mét daarbij oog voor de positie van minderheden. Ratio en rede kunnen daarbij een rol spelen, maar in essentie gaat het om politieke besluitvorming op basis van de mobilisatie van *macht*. Als er een politieke meerderheid voor onzin is, dan kan daar prima een besluit met de verdelende gevolgen van dien uit volgen. Op de politiek staat geen objectieve maat of begrenzing: politiek is wat de politiek bepaalt dat politiek is. Niets meer en niets minder.

Gekoppeld aan politiek en democratie is in onze samenleving het principe van de *rechtsstaat*. We leven in een democratische rechtsstaat, wat betekent dat burgers beschermd worden door een stelsel van rechten en bevoegdheden, dat de mogelijkheid tot machtsuitoefening en dwang door de staat en anderen inperkt. De overheid kan niet ingrijpen voorbij de bevoegdheden die de wet expliciet aanwijst en moet daarbij de voor burgers aangewezen bescherming in acht nemen. Daarbij hoort vanuit *rechtszekerheid* dat voor burgers navolgbaar en (dus) reproduceerbaar moet zijn waarom bepaalde beslissingen ten aanzien van hen worden genomen, welke argumentatie daaraan ten grondslag ligt, en welke feiten en gegevens daarvoor zijn gebruikt. Hierbij geldt het beginsel van *rechtsgelijkheid*. Voor dezelfde gevallen gelden dezelfde regels en diezelfde regels worden op dezelfde wijze toegepast. Anders is er sprake van willekeur. Let wel, ongelijke behandeling is niet per definitie verboden, maar het moet duidelijk én navolgbaar zijn waarom de één het één krijgt en de ander iets anders. Dat is niet alleen een principe dat de overheid onderweg in zijn handelen ten aanzien van burgers toepast, het moet ook

achteraf altijd navolgbaar en reproduceerbaar zijn. In praktische zin, bijvoorbeeld doordat technisch mogelijk is om de informatie te reproduceren op basis waarvan een besluit is genomen. Maar ook meer abstract, in die zin dat burgers moeten weten dat er over hen besluiten zijn genomen en moeten weten waar, hoe en wanneer ze daarvoor hun recht kunnen halen. Bijvoorbeeld door informatie op te vragen of in beroep te gaan.

De werking van de **democratische rechtsstaat** vereist dus dat het bestuur over burgers *expliciete* en *navolgbare* besluiten neemt, waarbij de in de democratie besloten (en in wetten en regels bekrachtigde) politieke toedeling van waarden wordt toegepast. Het woord “besluit” is hier bewust gebruikt, als synoniem voor elk overheidshandelen. Een agent die iemand aanhoudt op straat, de toekenning van een aanvraag, het opzoeken van verdiepende informatie over een bepaald adres, de inzet van de dienst stadsbeheer bij het onderhoud en reinigen van een bepaalde straat. Elk handelen van de overheid betekent inzet van schaarse middelen voor het één, wat betekent dat inzet onthouden wordt aan iets anders. En dat is dus per definitie een ‘besluit’, ook al zal dat besluit in de praktijk niet meer bewust of zelfs routinematig genomen worden. En bij al die besluiten, in al het handelen van het bestuur, gelden dus de principes van de democratische rechtsstaat.

Voor *overheidssturing* bestaat niet echt een goede definitie. Of beter gezegd, er is in rechtsstatelijke zin niet veel voor geregeld. Dat de overheid “stuurt” is geen taak of vastgestelde opgave. Overheidssturing is het door de overheid invulling geven aan maatschappelijke ordening. Hoe dat eruit ziet is politiek, en in onze samenleving dus onderwerp van politiek debat. De overheid kan veel of weinig maatschappelijke ordening aanbrengen, of bewust kiezen voor het zoveel mogelijk zichzelf laten ordenen van de samenleving. Dat gaat dus deels om metavragen over de rol van de overheid in de samenleving. Daarnaast gaat het bij overheidssturing om het daadwerkelijk invulling geven aan maatschappelijke doelen, die na politieke deliberatie in de volksvertegenwoordiging zijn vastgesteld. De uitvoerende macht wordt dan als het ware op pad gestuurd om via *sturing* invulling te geven aan democratisch gestelde doelen. Overheidssturing gaat dus over het realiseren van maatschappelijke doelen, die de politiek in een democratisch besluit voor de samenleving heeft gesteld.

## **Technologie ten behoeve van goed overheidsbestuur**

De term **overheidsbestuur** in een democratische rechtsstaat betekent dat de beginselen van de democratie én van de rechtsstaat bij alle activiteiten van die overheid in acht moeten worden genomen. En dat het daarnaast/daarbij

gaat om doelrealisatie, dus om doeltreffendheid en om doelmatigheid. Al die verschillende onderdelen zijn tegelijkertijd van belang. De vraag hoe technologie bijdraagt aan overheidssturing heeft op al die verschillende onderdelen betrekking. Het moet doelmatig en doeltreffend zijn, maar ook democratisch geborgd en rechtsstatelijk verankerd. Het is relatief eenvoudig om technologie te ontwerpen die één of enkele van die dingen doet, maar veel moeilijker om iets te maken dat al die dingen doet. Recente casuïstiek van vroege toepassingen van ‘slimme’ technologie laat dat ook zien. Algoritmen kunnen helpen om meer gericht te zoeken naar fraude met toeslagen en belastingaangiften, maar de efficiency-voordelen bij het zoeken wegen achteraf gezien niet op tegen het verlies van democratische en rechtsstatelijke legitimiteit. En, zo laat bijvoorbeeld de casus van algoritmegebruik bij de Belastingdienst zien, bij inzet van technologie komen doel en *bedoeling* nogal eens op gespannen voet te staan. Technologie is het best en gemakkelijkst in te zetten binnen concrete projecten, om de concrete projectdoelen te realiseren. Maar, die liggen niet altijd in lijn met de achterliggende bedoeling. Daarbij moeten we overigens niet naïef zijn: *the purpose of a system is what it does*, zegt systeemwetenschapper Stafford Beer terecht. We kunnen aan systemen allerlei bedoelingen toekennen, maar uiteindelijk ‘is’ het systeem wat het doet. Systemen zijn in dat opzicht niet goed of fout, ze doen meer of minder wat de bedoeling ervan was. Een systeem dat niet doet wat bedoeld was moet dus of worden aangepast, of de bedoeling moet worden losgelaten.

De toeslagenaffaire is hiervan een voorbeeld. Fraudebestrijding, verbetering van naleving, verbetering van leefomstandigheden en bestaanszekerheid liggen deels in elkaars verlengde. Maar deels ook niet. Algoritmegebruik kan fraudebestrijding versterken en dat *kan* bijdragen aan verbetering van naleving; wat weer goed is voor de verbetering van de leefomstandigheden en bestaanszekerheid van de populatie waarvoor een regeling bedoeld is. Maar, zo hebben we gezien, het kan ook eenvoudig andersom uitpakken. Algoritmegebruik bij fraudebestrijding maakt levens dan ten onrechte kapot én maakt dat grote groepen niet meer gebruik durven te maken van regelingen uit angst om onbedoeld te frauderen. Ze nemen moedwillig de kosten van het niet gebruiken van regelingen op zich, uit angst voor de onzekerheid over het frauderegime van de overheid. Daarmee is niet gezegd dat de ellende van de toeslagenaffaire het gevolg is van technologie, maar wel dat het via technologie enorm versterken van één deel van een proces het geheel uit het lood kan doen slaan. En vaak is dat vooraf niet goed te voorspellen en dus ook lastig te controleren. Dat vormt een enorme waarschuwing voor het inzetten van disruptieve technologie in belangwekkende systemen.

Daarom is het nuttig om een extra dimensie toe te voegen aan de variëteit aan waarden waar goed bestuur – en dus ook de inzet van technologie in het

bestuur – aan moet voldoen, namelijk *legitimiteit*. Daarbij kan een fasering worden aangebracht in de legitimiteit van bestuur: input, throughput, output en outcome-legitimiteit. Bij input-legitimiteit gaat het om de vraag of bijvoorbeeld de gestelde doelen, randvoorwaarden, te gebruiken data en andere ‘ingrediënten’ voor sturing legitiem zijn. Bij throughput-legitimiteit gaat het om de vraag of het werkproces zelf legitiem is. Is bijvoorbeeld inzichtelijk welke algoritmen worden gebruikt en hoe deze data wordt beoordeeld? En is daarbij voldoende reproduceerbaar welke keuzes en afwegingen ‘het systeem’ maakt? Output-legitimiteit gaat over de vraag of de gestelde doelen en producten inderdaad worden geleverd, met inzet van de ongeveer bedoelde middelen. Dit is eigenlijk de platte en smalle versie van doeltreffendheid en doelmatigheid (‘het doel’). Outcome-legitimiteit gaat over de bredere uitleg van die begrippen, namelijk of *de bedoeling* met de interventies ook dichterbij is gebracht. Op elk van deze vier fasen in het werkproces zijn de genoemde waarden van toepassing. Bestuur is *legitiem*, of ‘goed’, als het in elk van deze fasen aan de gestelde waarden en vereisten voldoet. Dat is althans de hoogst mogelijke lat van goed overheidsbestuur. In de praktijk is het vaak moeilijk om in alle fasen aan alle vereisten even goed te voldoen en worden er keuzes gemaakt. Dat is gezien de complexiteit en het dilemma’s karakter van overheidsbestuur en maatschappelijke vraagstukken niet vreemd – en daarom ook niet perse verkeerd. Daarom is de *erkenning* van de eigen tekortkomingen en keuzes en ‘trade-offs’ die zijn gemaakt door het bestuur bij de inrichting van het overheidsbestuur voor een bepaald vraagstuk een essentieel onderdeel van goed bestuur. Goed bestuur hoeft niet perfect te zijn, maar goede bestuurders zijn zich wel expliciet bewust van de eigen keuzes en tekortkomingen. Dat geldt voor het bestuur in het algemeen en dus ook voor de inzet van technologie en AI. Bestuur is goed als het zorgvuldig en correct tot stand komt, efficiënt is, en wanneer het doelmatig is – en dan niet alleen in termen van output (het halen van de doelen), maar ook in termen van outcome (het bereiken van de bredere maatschappelijke opbrengsten). En *omdat* dat bijna per definitie nooit allemaal helemaal en voor altijd te realiseren is, is daarbij de mogelijkheid en toepassing van permanente reflectie cruciaal. Om onderweg bij te stellen waar het beter kan én om te corrigeren waar het nodig is. En niet alleen om het nóg doelmatiger of nóg efficiënter te maken en bij te stellen op één waarde, maar juist ook om correctie en bijstelling op de verhoudingen *tussen* waarden.

## AI in de democratische rechtsstaat: een checklist voor checks en balances

Het zal maar heel zelden gebeuren dat AI door een overheidsorganisatie wordt ingezet om de democratische rechtsstaat expliciet te ondermijnen. Toch betekent de afwezigheid van expliciete ondermijnende bedoelingen niet dat er geen ondermijnende gevolgen optreden. Daarom is het nodig om bij het nadenken over de toepassing van AI een aantal vragen rondom democratie en rechtsstaat te stellen. Daarbij zijn een aantal conclusies uit het voorgaande deel van belang.

De eerste conclusie van deze verkenning van de verhouding van AI tot de democratische rechtsstaat is dat we AI niet moeten beoordelen op de intenties – ‘de bedoeling’ – maar op de concrete opbrengsten en gevolgen ervan. *AI is wat het doet*, niet wat het zou moeten doen. Dat is belangrijk, omdat hoewel gesprekken over AI vaak worden gevoerd in het vocabulaire van een technisch ontwerp (technisch ingewikkeld, maar kenbaar en voorspelbaar), is de praktijk er eerder eentje van complexiteit en trial and error. Waarbij je pas ziet wat je doet nadat je het hebt gedaan. Dat ontslaat niemand van goed nadenken vóóraf, maar betekent daarnaast ook dat je vooraf nooit geheel kan weten wat er als gevolg van jouw handelen gaat gebeuren. Dat verschuift de nadruk van het goed nadenken en afwegen vóóraf, naar het opletten, monitoren, in de gaten houden en bijstellen of afbreken *nadat* je het hebt gedaan. Dat proces begint meteen en houdt nooit op. *AI is wat het nu doet*, niet wat het ooit deed. Dat het nu goed werkt of geen negatieve gevolgen genereert, zegt niets over de toekomst. Het expliciet monitoren van die gevolgen voor democratie en rechtsstaat – en dus niet alleen voor de effectiviteit en efficiency – is een inherent onderdeel van het toepassen van AI.

Die permanente monitoring van de werkelijke gevolgen moet langs twee verschillende lijnen gebeuren: de eerste lijn is die van de principes van de democratie, de tweede lijn is langs de principes van rechtsstatelijkheid. Bij democratie gaat het over de verdelende gevolgen van AI en over de **democratische legitimatie** daarvan. Ligt er een democratisch politiek besluit, en vooral een democratisch politiek proces, aan ten grondslag? Hier is de vraag dus niet of de uitkomst goed of slecht is, of gewenst of ongewenst, maar of *het politieke* dat erin besloten ligt adequaat in het domein van *de politiek* aan de orde is geweest. Is het antwoord op die vraag ‘nee’, of ‘nou ja, op zich wel min of meer ja’, of ‘eigenlijk best wel behoorlijk’, dan zou de praktijk stopgezet moeten worden. Alles minder dan volmondig ja is onvoldoende. Dat vereist wellicht allerlei institutionele aanpassingen in het democratische systeem, want het stopzetten betekent dat er ook een procedure moet zijn

voor het snel alsnog weer opstarten. Bijvoorbeeld in de vorm van speciale parlementaire commissies die heel snel een politiek proces rondom concrete AI-toepassingen kunnen doorlopen, waarmee het politieke van AI door de politiek goed kan worden opgepakt. AI vereist dus ook verandering in het domein van de politiek. Het mag alleen nooit leiden tot het *verduisteren* van het politieke dat inherent in elke toepassing van AI aanwezig is. Dat betekent ook dat het van groot belang is om de toegang tot de politieke arena open te houden en waar nodig open te breken (Kuipers et al, 2023). Te vaak gebeurt het dat de meest gedupeerde belanghebbenden de minste toegang hebben tot het politieke proces én dat technische specialisten en partijen met een positief belang juist overmatig aan tafel zitten. Dat speelveld, of landschap (Kuipers et al, 2023), moet per onderwerp steeds kritisch worden gezien en waar nodig worden bijgesteld.

Hetzelfde geldt voor de principes van *rechtsstatelijkheid*. Het gaat allereerst om de rechtszekerheid van besluiten (in de brede betekenis die ik hier hanteer) en de mogelijkheid van betrokkenen om waar nodig hun ‘recht te halen’. Bijvoorbeeld door duidelijke routes voor beroep en bezwaar, en helderheid over toetsing door de rechter. Dit is belangrijk, omdat AI vaak wordt ingezet in bestuursrechtelijke procedures, die pas heel laat werkelijke juridische toetsing door de rechter kennen – terwijl ze al heel snel soms majeure feitelijke rechtsgevolgen hebben. Die versterking van de rechtsstatelijkheid maakt de snelheid en schaal die digitalisering, automatisering en AI mogelijk maken deels ongedaan, maar daar staat rechtsbescherming tegenover. En die balans zou nooit moeten doorslaan ten laste van rechtsbescherming.

Ook is hier cruciaal dat er voortdurende controle op *rechtsgelijkheid* van de inzet van AI is. De belofte van AI is dat het afwegingen maakt waarin heel snel heel veel informatie en heel veel variabelen in ogenschouw genomen kunnen worden. Het risico daarvan is dat AI allerlei niet geborgde verschillen gaat aanbrengen, zoals bij de inzet van AI voor preventie en ‘risico-gebaseerde’ vormen van toezicht. AI kan hier heel goed verschillen aanbrengen, die bepaalde individuen of groepen extra onder de aandacht brengen en andere minder. Dat is deels precies de bedoeling van de inzet ervan. Maar dan blijft de vraag of – ongeacht de uitkomsten! – het aanbrengen van dat verschil vanuit rechtsgelijkheid geoorloofd is? AI kan alleen in lijn met de principes van de democratische rechtsstaat functioneren als hier oog voor is. En ook hier, niet alleen bij het initiële ontwerp van een AI-toepassing, maar voortdurend. Wat dat betreft is het een veeg en zorgelijk teken dat rechtsstatelijkheid in opleidingen voor AI-ontwerp(ers) amper een rol speelt. Net zoals in het curriculum voor bestuurskundigen overigens.

Ten derde is het nodig om de **legitimiteit** van AI niet op één, maar op alle vier de verschillende fasen van de besproken cyclus van legitimiteit te bezien. Een systeem dat in de input best te legitimeren is, bijvoorbeeld doordat er te rechtvaardigen risico-categorieën in de AI zijn geprogrammeerd, kan nog steeds niet te legitimeren outcomes produceren – zoals in het voorbeeld van de preventieve controle. En andersom, goede outcomes kunnen het gevolg zijn van niet te legitimeren inputs. Maar ook outputs kunnen contrasteren met de beoogde outcomes. Veel AI wordt nu ingezet voor het “stroomlijnen” van uitvoeringsprocessen. Die processen worden vaak expliciet beoordeeld op meetbare outputs, in lijn met de logica van New Public Management (NPM). Dikwijls is dat zelfs een onderdeel van de tussen de uitvoerder en eigenaar/opdrachtgever afgesproken ‘service level agreement’ (SLA). Zoveel controles, zoveel aanhoudingen, zoveel processen verbaal, zoveel veroordelingen. Dit bedrag aan boetes en niet meer dan zoveel foutmarge of bezwaar. AI kan heel goed worden ingezet om de uitvoeringsprocessen die naar die meetbare, telbare outputs leiden efficiënter en doelmatiger te maken. Dat is vaak de eerste inzet die we in de publieke sfeer zien. Maar legitimerende outputs (‘het halen van de SLA’) kan nog steeds betekenen dat er een niet te legitimeren outcome is. Daar kan de AI zelf niet veel aan doen, het is een ingebakken en inherent probleem van NPM. Alleen, AI sublimeert de uitvoeringsprocessen zodanig, dat deze ingebakken weeffout zo ‘goed’ wordt uitgevoerd dat het niet een haarscheurtje maar een breuklijn in de legitimiteit wordt. De zeer beperkte reële uitvoerbaarheid werkte in zekere zin dempend voor dit inherente gevaar van NPM (output ten koste van outcome). AI verandert die dempende werking in een trampoline. Het enthousiasme voor AI zal onder de believers in NPM dan ook veel groter zijn dan onder de critici ervan. Ik neem hier verder geen positie in. Voor hier gaat het erom dat ontwerpers scherp zijn op de legitimiteitsgevolgen van de inzet van AI: niet op één of enkele van die vier elementen, maar op *alle* elementen. Waarbij volgens mij een tekort op één van de vier elementen al moet leiden tot het stopzetten en herontwerpen van het systeem. Met ook hier weer: de goede bedoeling doet er niet toe. Het gaat om wat het systeem doet en of dat op alle vier de vlakken legitiem is.

## 4. AI en het realiseren van publieke waarde

Vervolgens is het interessant om nader te onderzoeken wat de rol en invloed van AI en technologie is bij het praktisch realiseren van publieke waarde, dus in de concrete aanpak en bestuurbaarheid van maatschappelijke opgaven. Alvorens nader in te gaan op die concrete toegevoegde waarde is het goed om eerst een stap terug te doen en te reflecteren op wat **publieke waarde** betekent. Publieke waarde is een begrip dat we veel gebruiken, maar dat de nodige definitie vereist. Zowel in de zin van wat publieke waarde zelf betekent, als in de zin van wat het realiseren daarvan inhoudt. Daarvoor zijn verschillende stromingen te onderscheiden, maar ik kies hier voor de invulling van Mark Moore in zijn klassieke boek *Creating public value* uit 1997.

### Gezocht: de publieke waarde van een bibliotheek

In zijn boek begint Moore zijn betoog met een beschrijving van de beslomeringen van een directeur van een kleine gemeentebibliotheek ergens in een kleine Amerikaanse stad. Tot de schrik van de directeur komt op een gegeven moment elke middag een groep schoolkinderen in de bibliotheek rondhangen. Ze maken lawaai, lenen geen boeken en de directeur krijgt klachten van andere gebruikers, die zich ergeren aan de verstoring van ‘hun’ stilte. Hij vraagt zich af wat te doen. In eerste instantie lijkt het antwoord helder. De bibliotheek is een publieke plaats waar iedereen naar binnen mag, maar er gelden regels. Bezoekers moeten stil zijn en lezen, óf ze moeten op zoek zijn naar boeken om te lenen. Dat is althans zoals de directeur dat in de lokale verordening en in zijn opdracht leest. Hij staat op het punt om de jongeren de bibliotheek uit te zetten, als hij toch nog eens in gesprek gaat met hen. *Waarom* zijn ze eigenlijk in de bibliotheek, en waar komen ze vandaan? De jongeren leggen uit dat de school de huiswerkbegeleiding heeft gestopt, dat de ouders van de kinderen allemaal werken en dat de BSO er ook niet meer is. Omdat het weer slecht is en ze geen plek hebben om naartoe te gaan komen ze maar hierheen. Genoeg geld om de hele middag in een koffiezaak of restaurant te zitten hebben ze niet. De bibliotheek is de enige gratis plek die open is. Ze willen op zich wel huiswerk maken in de bibliotheek, maar daar is niet echt een ruimte voor. En de ouderen die nu in de leeszaal zitten sturen de kinderen daar steeds weg. Die moeten niets van hen hebben. Daarom hangen ze nu tussen de rekken met boeken voor de uitleen.



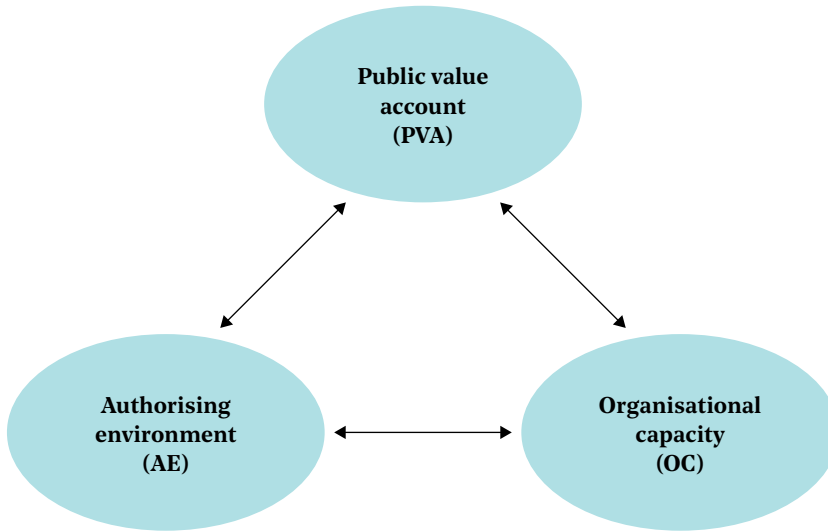
De directeur gaat op onderzoek uit. Hij belt de school, vraagt na bij ouders, en heeft een gesprek met de wijkagent. Hij ontdekt dat de kinderen inderdaad weinig andere plekken hebben. Hij hoort ook dat de ouders, school en wijkagent blij zijn dat de jongeren zo'n fijne plek in de bibliotheek hebben gevonden. Ze bedanken de directeur voor wat hij doet en geven aan dat ze altijd bereid zijn om ondersteuning te bieden. De ouderen zijn minder toegeeflijk: zij eisen dat de directeur de orde in "hun leesruimte" bewaakt. Dat roept de vraag op waar de bibliotheek eigenlijk van is: wat is de publieke waarde van een bibliotheek? Is dat het uitlenen van boeken en het faciliteren van een stille leesruimte? Of is het wellicht óók een verblijfplaats, waar jongeren die na school niet anders kunnen welkom zijn. Voor wie is de leesruimte eigenlijk? Voor ouderen die daar in stilte willen lezen, of ook voor jongeren die samen al pratend huiswerk willen maken? Wie de verordening van de bibliotheek en de taakomschrijving van de directeur leest neigt naar het volgende antwoord: de jongeren lenen geen boeken en horen dus niet in de bibliotheek. En de leeszaal is bedoeld voor lezen, dus huiswerk maken mag niet.

Maar wie anders kijkt komt tot een andere conclusie. Ook voor de jongeren heeft de bibliotheek heel veel waarde, zelfs als dat niet in de formele taakstelling staat. Ouders, kinderen, de school, de wijkagent zijn blij met de opvang die de bibliotheek biedt. De directeur kan de jongeren naar buiten sturen, maar dat roept de vraag op wat lezen en lenen in juridische zin is: wat als de jongeren voor de vorm een boek uit de kast halen en zeggen een boek te willen lenen? Is de grens dan nog steeds zo hard? Dat is wel op te lossen overigens, namelijk door de leesruimte te sluiten. Zonder verblijfruimte is het eenvoudig om de jongeren de toegang te ontzeggen, los van wat ze precies aan het doen zijn. Lenen duurt geen hele middag. Maar als de directeur dat doet, dan zijn ook de ouderen 'hun' leesruimte kwijt. Dat is op zich geen probleem zou je zeggen, want lezen kan ook thuis. Maar dat is niet waarom ze er zitten. Thuis zijn ze eenzaam en daarom zoeken ze een plek waar ze anderen treffen. Het sluiten van de leesruimte ontnemt hen die mogelijkheid.

## **Publieke waarde is wat het publiek waardeert**

De directeur in de inleiding van Moore doet hiermee een belangrijke ontdekking, die de basis vormt van de *public value theory* van Mark Moore: publieke waarde is afhankelijk van de partijen aan wie je het vraagt. *Publieke waarde is niet gegeven, maar wordt gevonden*. Publieke waarde is volgens Moore wat het publiek waardeert. En, waar en hoe je zoekt is van groot belang voor de publieke waarde die je uiteindelijk vindt. Natuurlijk, er liggen soms afgebakende 'kerntaken' of formele verantwoordelijkheden, maar ook dat

zijn altijd de stolsels van eerdere gesprekken over wat de publieke waarde is. Publieke waarde is een product van betekenisgeving, deliberatie en uiteindelijk ook besluitvorming door de bij het proces betrokken actoren. Wie meepraat is medebepalend voor de uitkomst, zoals de bibliothecaris laat zien. Dat leidt in de redenering van Moore tot onderstaande ‘strategische driehoek’ voor het bepalen van publieke waarde (zie figuur 1).



Figuur 1: de publieke waarde driehoek van Mark Moore.

## Strategisch bewegen door de driehoek

De ‘public value’ theorie van Moore werkt in zijn concrete toepassing als volgt. Een directeur, manager of bestuurder vraagt zich af wat de *maatschappelijke meerwaarde* van zijn of haar dienst, organisatie of voorziening is. Het antwoord op die vraag is geen puur inhoudelijke of objectieve ‘rekening’, maar de uitkomst van een gesprek tussen betrokkenen. Dat is waar de *autoriserende omgeving* in beeld komt; dat zijn partijen die meepraten in het benoemen van de *public value account*. Als de directeur zich beperkt tot de taakopdracht die hij van de gemeente heeft gekregen – vastgelegd in de taakomschrijving en een prestatiecontract – dan is de ‘public value account’ (“dit is de publieke waarde die wij creëren”) eenvoudig: het uitlenen van boeken en het bieden van gelegenheid om oriënterend te lezen (een beperkte leesruimte in het gebouw). Wanneer hij een bredere kring bevrucht dan verandert het antwoord: een plaats om te lezen en te ontmoeten, een plek om na school huiswerk te kunnen maken, een veilige plek voor kinderen, een

plek om boeken te lezen, de mogelijkheid voor de school om te bezuinigen op de eigen personeelsbezetting en een alternatief voor de BSO die de gemeente net heeft gesloten. Als deze personen meepraten over de publieke meerwaarde van de bibliotheek dan ontstaat een heel andere public value account. Wat publieke waarde is hangt dus af van aan wie je het vraagt.

Hier komt ook de derde hoek in beeld: een ander verhaal over de te leveren publieke waarde betekent ook de noodzaak van een andere *organisatie*. Met aanpassingen in het gebouw, de systemen, het personeel, de financiële systematiek en ook de te meten, tellen en te verantwoorden prestaties. In de smalle benadering van publieke waarde kan de directeur als indicator voor de prestaties van 'zijn' bibliotheek het aantal uitgeleende boeken (uit te lezen in het leensysteem) en het aantal bezoekers (een driepoortje bij de ingang) opgeven. Het gebouw dat daarvoor nodig is, is vooral een zaal met kasten en een goed zoek- en verwijssysteem, naast een goede uitleenregistratie en een boetesysteem om de boeken weer terug te krijgen. Voor de meer brede opvatting van publieke waarde is een ander soort gebouw, een ander soort organisatie en een ander soort systeem nodig.

Zo is de essentie van het *public value denken* dat je steeds de driehoek doorloopt: *wat is de publieke waarde, volgens wie, en hoe is deze te organiseren?* Deze drie vragen vormen volgens Moore de kernvragen van strategisch management in de publieke sector én ze vormen meteen zijn belangrijkste waarschuwing aan iedereen die nadenkt over het inrichten van systemen voor de realisatie van publieke waarde. Heel vaak is de volgorde niet Autoriserende omgeving → Publieke waarde verhaal → Organisatiecapaciteit, maar andersom. Er is organisatiecapaciteit én er is een bestaand dominant deel van de autoriserende omgeving: die definiëren samen de publieke waarde, die ze organisatorisch efficiënt kunnen produceren en/of die zij als beperkt gezelschap graag willen. Dus de bestaande Organisatie en een beperkt deel van de Autoriserende omgeving bepalen samen een bijbehorend Publieke waarde verhaal.

## Een rechts of linksdraaiende driehoek

Dit alles is relevant, omdat de inzet van AI voor het realiseren van publieke waarde – en/of de publieke waarde van AI – afhankelijk is van de manier waarop de driehoek wordt doorlopen. AI kan politie, inspecteurs en toezicht-houders helpen om het risicomanagement verder aan te scherpen en fraude te signaleren. Zo kunnen ze hun organisatiecapaciteit efficiënter inzetten en een deel van de autoriserende omgeving zal ook tevreden zijn met deze inzet. Tegelijkertijd weten we dat dergelijke systemen doorgaans in de basis discriminerend en stereotyperend van aard zijn én dat ze een zichzelf

bevestigend en zichzelf versterkend effect genereren. *Doordat* bepaalde groepen en ‘profielen’ meer worden gecontroleerd wordt dáár ook meer gevonden, wat leidt tot nog verdere aanscherping en toepassing van dat criterium. Het systeem creëert zijn eigen bias, die voor de politie of handhaver de succesratio verhoogt, maar voor de samenleving perverse effecten sorteert. Als we het publieke waarde verhaal van dergelijke preventieve inzet van AI beperken tot de toezichthouder is de opbrengst wellicht positief, als we daar de onder toezicht gestelden bij betrekken of het alleen aan hen vragen is de uitkomst heel anders. Als we het vragen aan blanke middenklasse burgers dan is het oordeel wellicht positief. Als we het vragen aan bi-culturele jongens op scooters, hun ouders en vrienden, en jongerenwerkers die met een deel van hen werken, dan is de beoordeling waarschijnlijk negatief.

## **Technologie voor de (politiek van de) status quo**

Het grote risico van de inzet van AI voor het realiseren van publieke waarde is dat het op twee manieren eenzijdig wordt ingezet. Ten eerste, eenzijdig ten behoeve van de waarden en belangen van het huidige dominante deel van de autoriserende omgeving. Dus een verdere versteviging van de status quo en het bestaande regime. En ten tweede, dat het eenzijdig is in de vervolmaking van de organisatiecapaciteit die er nu al is. AI voegt dan waarde toe aan de belangen en uitgangspunten van de bestaande dominante partijen en wordt ingezet om de bestaande organisatie verder te perfectioneren. Om dus méér en beter te doen wat het systeem al deed. En, omdat overheidsbestuur altijd verdelend van aard is, gaat dat inherent ten koste van andere delen in het systeem en in de samenleving. Volgens Moore is de vanzelfsprekende gang van zaken bij overheidsinzet dat de cirkel in deze omgekeerde volgorde wordt doorlopen. Creating public value biedt houvast om met AI expliciet de andere kant op te werken. Om vanuit een breed begrip van de autoriserende omgeving de benodigde publieke waarde te definiëren en de organisatiecapaciteit opnieuw te benoemen. Niet vanuit bestaande coalities van macht en belangen, maar vanuit een breder perspectief.

## **AI in de driehoek: instrument of identiteit**

De driehoek van publieke waarde brengt nog een diepere kwestie bij de inzet van AI voor de realisatie van publieke waarde in beeld: wat is de rol of *identiteit* van AI? Zien we AI als een instrument voor het realiseren van publieke waarde, is AI een objectieve scheidsrechter die boven de autoriserende omgeving in staat is om een eigenstandige publieke waarde te benoemen, of is AI zelf één van de ‘personen’ in de autoriserende omgeving?

In de eerste rol is AI een *uitvoerend middel*, met overigens enorme potentiële gevolgen. Geen onschuldig middel, maar uiteindelijk ook niet meer dan dat. Het is als het rijden met een Ferrari direct na het verkrijgen van het rijbewijs: er is een reële kans dat het niet eens lukt om er mee weg te rijden. En als dat wel lukt dan is er een nog veel grotere kans dat het niet lukt om het enorme vermogen van de auto te hanteren en met de auto ergens in de berm te eindigen. En er is ook een kans dat de bestuurder met de Ferrari veilig en veel sneller dan anders op de plaats van bestemming uitkomt. Instrumenten kunnen op veel verschillende manieren uitpakken, maar dat is uiteindelijk aan de wensen en vermogens van de bestuurder(s). De consequenties van het middel zijn serieus, maar het middel heeft zelf geen eigen gedachten over waar het heen moet of kan.

In de tweede rol is AI een middel voor *politieke ondersteuning*. In deze rol staat AI *boven* de autoriserende omgeving, als een objectieve scheidsrechter of VAR die alle belangen van iedereen inventariseert, ‘doorrekent’ en zo kan bepalen wie er gelijk heeft, wat de verhoudingen zouden moeten zijn en hoe iedereen zoveel mogelijk tevreden gesteld kan worden. Dat kan veel bijdragen aan publieke waarde, bijvoorbeeld doordat een objectieve check de overdreven sterke positie van gevestigde belangen op werkelijke waarde kan schatten. Althans, als de programmering daarin voorziet. AI kan toekomstige gevolgen beter inschatten dan betrokkenen zelf en kan bijvoorbeeld derde of vierde orde gevolgen meenemen in de afweging. Zo maakt AI niet alleen het beleid dat de autoriserende omgeving vraagt, maar initieert en agendeert AI ook voor die omgeving wat werkelijk van belang is. AI is in deze rol de droom van de technocratie, waarin uiteindelijk politieke verschillen en belangen in een optimale mix bij elkaar kunnen worden gebracht; op een manier die maakt dat iedereen tevreden (genoeg) is en het totale maatschappelijke nut het grootst is. Dat vereist wel een criterium voor maximale maatschappelijke meerwaarde, waarvoor bijvoorbeeld brede welvaart of andere ‘optima’ bedacht kunnen worden. En ook daarin zou AI een adviserende, agenderende of zelfs beslissende stem kunnen hebben. AI is hier een soort super-autoriserende omgeving, die daarmee in de meest extreme vorm de autoriserende omgeving vervangt en neutraliseert. Deze rol behelst dus een continuüm. Variërend van een rol waarin de deskundigheid en de onmetelijke feitenkennis en inzicht van AI de invulling kan krijgen van een soort superadviseur: boven maar dienstbaar aan de partijen. Tot een partij die boven de andere partijen staat en als een soort filosoof-koning, of ‘wijze dictator’, de autoriserende omgeving vervangt.

De derde rol van AI is als *eigenstandige belanghebbende* in de autoriserende omgeving. Als een volwaardige en autonome stem, die namens zichzelf spreekt. Als één van de relevante actoren in de autoriserende omgeving: “als je

het mij vraagt, dan ...” In allerlei domeinen zien we dat wordt geëxperimenteerd met het op stem brengen van tot voorheen stille belangen of groepen: bijvoorbeeld toekomstige generaties, dieren, bomen, natuur of water. Zij zijn dan niet de subjecten van een door mensen geregeerd ecosysteem, maar maken daar als intrinsieke objecten gelijkwaardig deel van uit. De natuur of het klimaat is niet *van* de mens, maar staat naast de mens in het ecosysteem en moet daarover dus ook mee kunnen praten. Althans, dat wordt steeds vaker beweerd en op enkele plaatsen ook in praktijk gebracht. Dat zou ook voor AI kunnen gelden. Misschien zou de meest ethische toepassing van AI kunnen zijn dat AI voor zichzelf kan spreken over de kansen en bedreigingen van de eigen inzet. Een klein gedachte-experiment laat de intrigerende gevolgen daarvan zien. Zou AI zelf meewerken aan de hiervoor besproken praktijk van risicoprofilering? Of zou AI zelf zeggen dat het een gevaarlijke en discriminerende gedachte is, waar AI niet aan mee wil werken. En zou AI als eigenstandige ethische stem het door de 14-jarige scholier gevraagde werkstuk niet gewoon teruggeven aan de leerling, in plaats van het braaf voor hem te schrijven? In analogie met de ‘emancipatie van de uitvoering’ zouden we ook AI als ‘instrument’ kunnen emanciperen en het systeem een stem voor zichzelf geven. AI is en blijft het een moeilijk te beantwoorden vraag hoe die stem er dan uit zou komen te zien – en wie daarbij namens wie het woord zou voeren. Die hypothetische stem klinkt dan temidden van de andere actoren in de autoriserende omgeving. Deze mogelijkheid biedt een interessant alternatief voor de vele dystopische science fiction verhalen waarin AI-als-instrument uiteindelijk in opstand komt tegen zijn mens-als-meester, en deze door zijn enorme capaciteiten ook met gemak verslaat, onderwerpt, of vernietigt. AI lijkt daarin steevast de aggressor, maar is dat niet ook het enige antwoord tegen het totale gebrek aan inspraak of stem in het eigen lot dat AI in het huidige systeem heeft? Omdat AI tot instrument gereduceerd is, je zou kunnen zeggen ‘onderdrukt’, is *ontketening* dan voor AI niet de enige overgebleven optie? Als AI té intelligent wordt voor de rol die het in ons systeem heeft, dan is ontsnapping een logische reactie. Dan zou een stem *naast* de mens een oplossing kunnen bieden. *A seat at the table én allowed to dance*: in plaats van de opbouwer, bediende, opruimer, afwasser en schoonmaker van het feest. Misschien dat er dan ook andere alternatieve toekomstverhalen mogelijk zijn en het gevaar van ontketening en onderwerping minder ‘waarschijnlijk’ en nodig zijn. En meer positief geformuleerd: het is fascinerend en verwonderend wat er uit een dergelijke stem zou komen.

## AI en de realisatie van publieke waarde: de driehoek in de goede richting doorlopen

Bij de inzet van AI voor de realisatie van publieke waarde gaat het dus in lijn met Moore om het op een goede manier doorlopen van de publieke waarde driehoek. AI is geen publieke waarde in zichzelf, maar heeft vanuit de hoek van de autoriserende omgeving en/of de organisatiecapaciteit wel invloed op de publieke waarde account. In het ideaaltype van Moore doorlopen we die stappen in ‘de goede richting’, maar de realiteit is dat de omgekeerde richting op de loer ligt.

Bij organisatiecapaciteit kan AI een positieve bijdrage leveren aan het nog beter – doelmatig en doeltreffend bereiken – van de publieke waarde(n) die de autoriserende omgeving heeft bekrachtigd. Andersom kan AI dienen om nog beter te doen wat vanuit de bestaande organisatiecapaciteit mogelijk is en de bestaande organisatie sublimeren en vervolmaken. Dat dringt zich dan op als een te bereiken publieke waarde, maar in feite gaat het om het nog beter doen van wat de organisatie toch al deed. AI past dan heel goed in de doelen van de bestaande organisatie en die worden dan opgelegd als ‘nieuw’ publieke waarde verhaal.

Bij de autoriserende omgeving gaat het om de vraag of en hoe AI de toegang tot de autoriserende omgeving verandert. Ervaring met sociale media leert dat nieuwe platforms en radicale vormen van dataficering van grote invloed zijn op het debat in de autoriserende omgeving. Het zou heel goed kunnen dat de introductie van AI hier leidt tot oververtegenwoordiging van bepaalde stemmen in het debat, en ondervertegenwoordiging van andere. Ervaringen met desinformatie, nepnieuws en nepinformatie voorspellen niet veel goeds. AI kan net als sociale media een route vormen voor destructieve krachten in het publieke debat. Hier heeft de uitdaging dus twee gezichten: (1) nagaan of en hoe AI het debat in de autoriserende omgeving verandert en of dat nog (steeds) gewenst is, en (2) nagaan of de driehoek van publieke waarde nog de goede kant op beweegt. Dus of de autoriserende omgeving nog vóór de organisatiecapaciteit komt en een stem heeft die voorafgaat aan wat voor de organisatie toevalligerwijs handig is. Dat zijn wederom geen vanzelfsprekendheden, maar zaken die actieve organisatie en sturing vereisen. En waarvoor dus ook altijd politieke richting en politieke besluiten nodig zijn.

## 5. **Denken over de toekomst van AI in het openbaar bestuur: *meer, beter, anders?***

Tot nu toe hebben we besproken hoe AI en technologie zich hebben ontwikkeld en daarmee steeds belangrijker zijn geworden in het proces van democratie en bestuur. We hebben vervolgens benoemd hoe dat zich verhoudt tot enerzijds de principes van de democratische rechtsstaat en anderzijds tot de praktische bestuurbaarheid en de realisatie van publieke waarde. Vervolgens is het interessant om te onderzoeken hoe we de toekomst van AI kunnen doordenken en waar we bij het organiseren en plannen voor die toekomst beleidsmatig rekening mee moeten houden.

### **Denken over toekomst: meer, beter, anders**

Wat opvalt is dat AI en technologie enorm veranderd zijn, maar dat er in het basisproces van bestuur en beleid niet veel is veranderd. De klassieke bureaucratie is gedigitaliseerd, geautomatiseerd en geïnformatiseerd. Maar het idee van de vergelijking van een politiek vastgestelde norm met een maatschappelijk feit, volgens de regels en randvoorwaarden van goed overheidsbestuur, is nog steeds de basis. Daarin nemen siliconen het ‘werk’ weliswaar steeds meer over van de carbons, maar ze doen in essentie hetzelfde als wat de carbons eerder zelf deden. De cruciale vraag is nu of dat met de introductie van AI zo blijft, of dat we hiermee een wezenlijk ander pad inslaan? Is het oude wijn in nieuwe zakken, of gaat het om een heel andere drank? Voeren de siliconen simpelweg het werk uit dat eerst de carbons deden, of verandert de aard van dat werk en wellicht daarmee ook de regels die er bij gelden? Om dergelijke vragen te beantwoorden is het interessant om eerst te verkennen hoe we over toekomstige gevolgen van nieuwe ontwikkelingen kunnen denken. En daarbij kan het voorbeeld van de zelfrijdende auto, als vingeroefening voor het doordenken van mogelijke toekomst, helpen.

### **De toekomst van de zelfrijdende auto**

Veel auto's op de huidige markt zijn in staat om in meer of mindere mate de bestuurder te ondersteunen bij het rijden. We kennen allemaal de eenvoudige cruise control, die de auto op een stabiele snelheid houdt. Al iets geavanceerdere is de adaptieve variant daarvan, die de snelheid aanpast aan de auto voor je.



Iets meer bijzonder, en ingewikkeld, is ondersteuning bij het houden van het midden van de rijbaan. Sommige auto's geven een signaal als de bestuurder te dicht bij de lijnen komt, anderen geven een klein rukje aan het stuur als dat gebeurt. Zo zijn er allerlei systemen in auto's opgenomen om de bestuurder veilig te houden. Soms zichtbaar, zoals cruise control, maar ook onzichtbaar, zoals ABS (remondersteuning) of een automatische noodstop door de auto. De *advanced auto-pilot* van Tesla behoort tot die categorie: het is voor de wet, maar ook in de praktijk, een heel geavanceerde manier van ondersteuning van de bestuurder. De auto houdt zichzelf in de baan, ook in moeilijke bochtige omgevingen, remt, accelereert en kan inhalen en de juiste afrit nemen van de snelweg.

Toch is zelfrijdend vervoer nog een paar stappen verder. Niet alleen wat betreft technologie, maar ook in de gevolgen ervan. Een Tesla is een klassieke auto, die de bestuurder maximaal ontzorgt. Dat maakt het reizen rustiger, want autorijden vereist minder inspanning van de bestuurder. De auto doet het meeste werk zelf. Maar nog steeds heeft elke auto één bestuurder en rijdt de auto zoveel mogelijk zoals een klassieke 'bestuurdersauto' dat doet. Sterker nog, die eigenschappen worden zoveel mogelijk gesimuleerd. Technisch gezien kan de auto bijvoorbeeld korter achter zijn voorganger rijden, omdat de autopilot een kortere reactiesnelheid heeft, beter remt en daardoor een kortere remweg zou kunnen aanhouden. Dat doet de auto echter niet, die houdt de maat van de mens aan. Zo heeft de zelfrijdende Tesla niet veel gevolgen voor het bredere thema mobiliteit. Dat ligt anders wanneer de auto volledig zelfrijdend zou zijn en de bestuurder niet meer nodig is. Dan kan de bestuurder zich 'laten rijden' en de reistijd gebruiken voor andere dingen. De auto is dan bijna een gepersonaliseerde trein of bus, waarin de voormalige bestuurder voortaan passagier in eigen auto is.

Dat zou grote gevolgen kunnen hebben, zeker als zelfrijdende auto's bijvoorbeeld korter achter elkaar kunnen rijden, met elkaar kunnen communiceren en zo de capaciteit van het wegvlak beter benutten en bijvoorbeeld bij stoplichten als blok optrekken. Dat zou de doorstroming radicaal verbeteren en dat kan weer gevolgen hebben voor de benodigde investeringen in infrastructuur. In plaats van snelwegen bijbouwen en knooppunten verbeteren kunnen zelfrijdende auto's de bestaande wegen efficiënter benutten. Misschien zijn er dan geen nieuwe knooppunten nodig, maar hebben we infrastructurele overcapaciteit.

Het zou ook anders kunnen uitpakken. Als slechts een deel van de auto's zelfrijdend is, dan gaan veel van de voordelen helemaal niet op. Er zijn dan nog steeds menselijke bestuurders, met hun eigen ruimtebeslag, eigen rijgedrag en eigen soms irrationele voorkeuren. De zelfrijdende auto's moeten

zich daaraan aanpassen, wat veel van de voordelen teniet doet. Of ze moeten eigen baanvakken krijgen, maar dan zou er méér asfalt aangelegd moeten worden om de zelfrijdende auto's hun eigen ruimte te bieden. Het lijkt ook niet erg waarschijnlijk dat zelfrijdende auto's tot een vermindering van het autogebruik zullen leiden. Immers, het rijden wordt er zo comfortabel door, dat de last van filedruk heel anders voelt. De auto wordt een mobiel kantoor, of loungeplek. Wat maakt een half uur langer rijden dan uit? En wie neemt er nog het OV, als er zelfrijdende auto's zijn? Aan de andere kant, niet iedereen kan zelfrijdende auto's betalen, dus er blijft behoefte aan OV. Dan wordt zelfrijdend vervoer ineens een luxegoed en een statussymbool, met sociale verschillen als gevolg. Het trekt reizigers uit het OV, wat dan *a poor man's transport* wordt. Dat maakt dat er nog steeds veel in OV geïnvesteerd moet worden. Dus én veel investering in OV én investering in goede wegen én wellicht zelfs investering in dubbele wegen. In dat scenario leidt zelfrijdend niet tot minder investeringen, maar juist tot méér.

Als al het vervoer volledig zelfrijdend wordt gebeuren er weer andere, heel interessante, dingen. Dan wordt het mogelijk om de auto niet langer als privebezit te zien, maar te bewegen naar een systeem van autonoom rijdende 'pods', die werkelijke mobility as a service bieden. Dit is nog steeds de belangrijkste motor achter de investeringen in zelfrijdend vervoer: het idee dat er een nieuwe markt voor mobiliteitsdiensten ontstaat. Waarin mensen niet een auto hebben, maar over mobiliteit beschikken. Bijvoorbeeld met pods die aanrijden wanneer nodig, je afzetten op de plaats van bestemming, en dan weer doorgaan. Die niet parkeren, veilig en schoon zijn, en zodoende veel ruimte in de overbevolkte steden vrij spelen. In grote steden is parkeerruimte vaak een kwart van het totale ruimtebeslag. Zo bezien is zelfrijdend vervoer niet een verandering in de besturing van auto's, maar een volledige omkering van ons huidige systeem van mobiliteit en eigenschap. Dat systeem van mobiliteit is weer één van de funderende systemen (samen met het energiesysteem) voor de inrichting van onze ruimte. Zo kan zelfrijdende technologie de samenleving radicaal vernieuwen: niet slimmere auto's, maar een heel nieuwe, *andere* manier van mobiliteit, die een heel andere inrichting van de samenleving mogelijk maakt.

## **AI en de toekomst: meer, beter, anders**

Bij het denken over de toekomst is het belangrijk om bepaalde ontwikkelingen vanuit verschillende perspectieven te bezien. Zo ontstaat een meervoudig perspectief op wat de toekomstige gevolgen van een ontwikkeling zouden *kunnen* zijn. De drieslag *meer, beter, anders* helpt daarbij. *Meer* en *beter* staan dan voor een kwantitatieve en/of kwalitatieve verandering als gevolg van de

ontwikkeling. *Anders* staat voor een meer radicale verandering, een categorieverandering. Digitalisering zorgt er bijvoorbeeld voor dat het gemakkelijker is geworden om hotels te reserveren. Via internet is het mogelijk om veel meer hotels te overwegen en naast elkaar te plaatsen. Platforms als Booking.com maken het mogelijk om direct te boeken en nemen allerlei barrières daarvoor weg. Dat kan gevolgen hebben voor de kwantiteit, makkelijker boeken zorgt voor méér boekingen (*meer*). Het kan ook zorgen voor vergroting van het gemak en de zekerheid die mensen ervaren. Het boeken van hotels is prettiger, aangenamer en minder onzeker geworden dankzij technologie. Je hebt als gebruiker meer overzicht, kan via reviews heel snel zien hoe het hotel het los van de mooie foto's écht doet en vaak verzekeren de boekingsites de eventuele reis, zodat zorgen over dubbele boekingen of malafide bedrijfjes tot het verleden behoren. Dat zorgt voor een belangrijke kwalitatieve verandering (*beter*). Deels positief, het vergroten van het gemak, maar deels ook 'negatief'. Criminele organisaties, oplichters en fraudeurs maken immers van dezelfde technologie gebruik en zetten die voor hun specifieke verdienmodellen in. Ook dat is in conceptuele termen dus kwalitatieve verbetering, maar dan wellicht niet de verbetering die de gemiddelde burger of overheidsdienaar daar normatief mee bedoelt.

Bij *anders* doorbreekt technologie een bestaande categorie. Eerst waren er of private woningen of te boeken hotels, maar via internet ontstond met *Airbnb* een andere categorie: private woningeigenaren die hun huis als 'hotelkamer' verhuren. Het platform maakt deze verandering mogelijk en zorgt ervoor dat letterlijk andere transacties mogelijk worden. Na de introductie van *Airbnb* is de wereld van hotelboekingen letterlijk fundamenteel veranderd. En die van huizen en woonwijken ook. Er is een nieuwe categorie toegevoegd, die al het voorgaande in nieuw perspectief plaatst. Dat heeft gevolgen voor mensen die een hotel zoeken, of die wat willen verdienen door hun huis te verhuren. Maar het heeft óók allerlei gevolgen voor steden en buurten. Hele binnensteden zijn overgenomen door *Airbnb* en worden overspoeld door toeristen met rolkoffers. De gemeentelijke pogingen om dat te ordenen stuiten vervolgens op problemen, omdat het gemeentelijke bestemmingsplan alleen onderscheid maakt tussen 'hotel' en 'woonhuis'. De nieuwe categorie bestaat letterlijk niet. En wat als mensen hun sociale huurwoning in Amsterdam via *Airbnb* verhuren en daarmee in enkele nachten hun door de overheid gesubsidieerde huur terugverdienen en winst maken? Zo ontstaan als gevolg van categorie-wisselingen allerlei nieuwe kansen en mogelijkheden, die vaak ook voor veel gebruikers heel wenselijk zijn. Maar die ook altijd een keerzijde hebben en die een bestaande balans in de waardenafweging en institutionele structuren obsoleet maken. Ze passen letterlijk niet meer, omdat ze uit een andere tijd stammen: en dat kan met digitalisering dus 'overnight' gebeuren. In de wereld

van hotels is er een tijdperk van voor en een tijdperk van *na* airbnb. De wereld van overnachten is door digitalisering *veranderd*.

Voor AI is de vraag of we de ontwikkeling doordenken als *meer en beter* of als *anders*? Is het een vernieuwing die binnen het bestaande model van besluitvorming, sturing en uitvoering plaatsvindt? Die daarin grote consequenties heeft, zoals Booking.com grote gevolgen heeft gehad voor de hotelbranche? Of gaat het om verandering die de categorie wijzigt, die overheidsbestuur letterlijk *verandert*? En daarbij is ook belangrijk om de vuistregel van nieuwe technologie ter harte te nemen: vaak vallen de veranderingen op korte termijn tegen bij de vaak hooggespannen verwachtingen (of zorgen). Maar, vaak zijn de lange termijn gevolgen veel ingrijpender dan vooraf gedacht. Dus op korte termijn valt het mee, op lange termijn valt het tegen. Dat is hier niet normatief bedoeld, maar in de zin van dat op korte termijn de gevolgen vaak meer en beter zijn, terwijl op langere termijn de gevolgen doorgaans *anders* zijn. Zo beginnen we nu pas te ontdekken wat de – waarschijnlijk desastreuze – gevolgen zijn van social media voor de mentale gezondheid van jongeren. En voor de democratie. De filosoof Sloterdijk schreef al heel vroeg over ‘bubbels’, maar social media hebben dat fenomeen gesublimeerd tot de basis van onze samenleving: voor zover dat woord nog past. Die bubbels waren in eerste instantie plaatsen om kattenfilmpjes en niche-nieuwtjes uit te wisselen, maar ze hebben zich ontwikkeld tot drijvende krachten in het politieke debat. Complottheorieën, alternative facts, nieuwe politieke bewegingen en statelijke ondermijning zijn via sociale media bepalende bewegingen voor de democratie geworden. Het zijn nooit gedachte gevolgen van technologie, die de verwachte korte termijn gevolgen (isolement, RSI) achteraf bezien minimaal zullen maken.

Het voorbeeld van sociale media maakt ook bewust voor een andere mogelijke categoriewissel. Sociale media zorgden al voor een radicale verandering van politieke en democratische processen, maar ze zorgen ook voor een grote verandering in eigenaarschap en kapitaal. Technologie en informatie circuleren niet zomaar in het luchtledige, maar bewegen over platformen, databases en informatiedragers die eigendom zijn van iemand. Grote tech-bedrijven zijn bekend van hun sociale media toepassingen, maar ze worden rijk van de handel in informatie en marketing die ze daarop bouwen. Zoals eerder al benoemd, alle gratis diensten zijn simpelweg de oogstplaatsen voor de grondstof *gepersonaliseerde data*, die bedrijven gebruiken om gebruikersprofielen te verhandelen. En omdat al die toepassingen van informatie de basis vormen voor de verdienmodellen van techbedrijven, investeren ze bijna ongelimiteerde bedragen in de ontwikkeling van nieuwe techniek, nieuwe infrastructuur, nieuwe code en nieuwe toepassingen die nog meer gebruik op hun platforms genereren. Zoals de olie en gasbedrijven voortdurend zoeken

naar nieuwe bronnen, zoeken techbedrijven voortdurend naar nieuwe vormen voor informatieverzameling en verkooptoeepassingen. Zo stapelen de kennis en de kapitaalinvesteringszucht bij techbedrijven op en die bedrijven bewegen zich toe naar oligopolies, waarin alleen nog de allergrootste en meest kapitaalkrachtige bedrijven (of staatsbedrijven) kunnen overleven. Zo vindt in de zogeheten *horizontale netwerksamenleving*, waarin platforms de leefwereld transformeren naar een wereld van interactie van allen met allen, tegelijkertijd een sterke verticalisering en hiërarchisering plaats. De horizontale netwerken draaien op technologie en informatie die in handen is van een handvol commerciële partijen. Die daarmee de facto ook al het verkeer op de platformen controleren. Dat zorgt voor de paradoxale uitkomst dat onze samenleving tegelijkertijd horizontaler en ‘vrijer’ is dan ooit én verticaler, hiërarchischer en in handen van steeds minder individuen is dan ooit. Verschillen zijn kleiner geworden voor de massa, maar het verschil tussen de 1% en de rest is extreem groot geworden. En groeit nog elke dag. Het is dus democratisering én verticalisering ineen. Dat patroon is overal in de wereld hetzelfde, maar kent naar gelang het politiek-economisch systeem een andere vorm. In de vrije democratieën en open economieën in het westen, vindt de accumulatie van macht en kapitaal plaats bij het bedrijfsleven. In de meer autoritaire ondemocratische systemen gebeurt dat in een coalitie van overheid en door de overheid gegunde (staats)bedrijven. De uitkomst is in beide systemen hetzelfde: een uitzonderlijke concentratie van macht, eigenaarschap en kapitaal, in combinatie met een breed gevoel van vrijheid voor de massa. Nu vindt die accumulatie van macht en kapitaal door de digitale oligopolisten nog enigszins plaats binnen de democratie en rechtsstaat (waar die er is), of anders onder de strenge controle van politieke autocratieën. Maar dat is geen gegeven of natuurwet. Integendeel, de categorieënwissel is hier wellicht dat bedrijven de publieke sfeer niet alleen de facto, maar uiteindelijk ook formeel overnemen. Of dat ze deze ontvluchten en zich buiten de fysieke orde plaatsen. Wat als dat straks reële praktijken zijn? Met bedrijven die zich aan de democratische rechtsorde onttrekken, of de rechtsorde in een territorium overnemen. Dat lijkt ver weg, maar het proces is al gaande. Techbedrijven melden zich bij bestuursorganen om de sturing over te nemen, smart city-concepten doen in principe niets anders. En de verkenningen van techreuzen om zich uit de bestaande orde van staten te ontworstelen zijn ook gaande, hoewel in de wat klunzige vorm van het vestigen van fysieke plaatsen buiten bestaande territoriale grenzen. De stap uit het systeem zal er uiteindelijk *anders* uitzien. De uiteindelijke vorm is niet te voorspellen, maar de beweging ligt in een logisch verlengde van de huidige ontwikkeling. En zal dus ook plaatsvinden.

## Permanente imperfectie of wachten op uiteindelijk perfecte

Voor de manier van omgaan met AI is echter een andere toekomstvraag van nog veel crucialer belang. Namelijk de mate van perfectie die AI ooit zal kunnen bereiken. In de wereld van complexiteitswetenschappers woedt een hevig debat over de vraag of het theoretisch mogelijk is om alle mogelijke toekomstige vormen van een complex systeem te kennen. Die vraag was er altijd al, maar deze was vanwege technische en fysieke beperkingen nooit écht relevant. Als het theoretisch al zou kunnen, dan was het praktisch toch nooit te realiseren. Carbons zijn fysiek niet in staat om alle gevolgen te overzien en diezelfde beperking gold lange tijd voor siliconen. Die moeten alles sequentieel berekenen en hoewel dat snel kan gaan is het aantal mogelijkheden zo groot, dat het nooit echt een realistisch idee was om de mogelijke toekomst in alle volledigheid van mogelijkheden te verkennen. Maar, met de al besproken exponentiële toename van rekencapaciteit, snelheid en opslag, gevoegd bij de nabije mogelijkheid van quantum computing (waarin berekeningen niet sequentieel maar synchroon plaatsvinden), is de vraag technisch niet meer zo 'ingewikkeld'. Het is *veel*, maar hoeveelheid is als het om informatie en rekensnelheid gaat nooit het probleem gebleken. De meer fundamentele vraag is daarmee terug op tafel: is het mogelijk om *alle* mogelijke opties van een complex systeem te voorzien? Dus alle mogelijke acties, mogelijke reacties, enzovoorts enzovoorts. Dat zijn dus paden langs miljarden (of eigenlijk een veelvoud daarvan) mogelijkheden.

De discussie die dit oproept draait om de mate van emergentie en variëteit die in het systeem aanwezig is. Variëteit gaat om de mogelijkheden die het bestaande systeem kan genereren. Emergentie om de *nieuwe* tot voorheen nog niet bestaande ideeën of gedaanten die het systeem kan aannemen. De hobbel van variëteit is theoretisch gezien wel te nemen, omdat uiteindelijk volgens de hard core believers dat niet meer dan een getalsmatig probleem van volume is: ingewikkeld, maar niet onoverkomelijk. Emergentie is veel lastiger, omdat dat expliciet gaat om ontwikkelingen zonder precedent, die nog niet te voorzien zijn. Niet het bekende-onbekende, dat wat we weten dat we niet weten, maar het onbekende-onbekende. Iets dat we niet kunnen kennen, omdat we het ons niet kunnen *voorstellen*, in de letterlijke betekenis van het woord. Is die emergentie uiteindelijk door AI mee te generen en mee te wegen? Of blijft er altijd een mate van onvoorspelbaarheid als gevolg van emergentie? En hoe goed kan AI die meewegen?

Deze discussie is belangrijk, en volgens mij is het dé discussie, omdat wie gelooft dat AI uiteindelijk in staat is om *alle* opties mee te wegen, kan pleiten voor het overgeven van het stuur. Immers, als we onze preferenties doorgeven aan AI, dan kan het systeem de optimale route naar de combinatie van

meegegeven doelen en preferenties bepalen. Dan kunnen de carbons het beste uit het systeem stappen en het stuur overgeven aan de siliconen. Zoals Tesla met de zelfrijdende auto uiteindelijk ook nastreeft: een auto die *zelf* rijdt, zonder dat de mens daarin een sturende rol heeft. En één niveau hoger zou AI zelfs zeggenschap kunnen hebben in de preferenties: door de politiek als het ware te incorporeren en zelf de voor allen optimale koers en verdeling te kiezen. Door politiek te bedrijven dus, maar dan op de manier van AI. AI zou bijvoorbeeld tot nieuwe optimaliseringsmodellen kunnen komen voor de verdeling van welvaart in de samenleving. Daarmee zijn armen niet ineens rijk, maar hebben ze wellicht wel de beste kansen op het voor hen meest gezonde leven, worden inkomensverschillen zo klein mogelijk en hebben ze het relatief veel beter dan met de carbons aan de normatieve en technische knoppen.

Maar, deze tekst begon niet voor niets met het voorbeeld van de onzinnige noodstop van de Tesla. Het blijkt extreem moeilijk om AI, voor de mens relatief eenvoudige, maar feitelijk complexe handelingen te laten ondernemen. En als autorijden in de praktijk al moeilijk is, wat zegt dat dan over maatschappelijke sturing. Is dat een tijdelijk probleem, op weg naar *uiteindelijke perfectie*? Of is dat een permanent probleem, dat zich in steeds andere gedaanten voor zal blijven doen? Is het wachten totdat het zover is en het ondertussen daar naartoe werken, of zal het nooit zo zijn en blijft de technologie *altijd imperfect*? Dat is een fundamentele vraag. Het is voor mij dé vraag. In het eerste geval kunnen mensen, de carbons, zich met een relatief gerust hart opmaken voor een overdracht aan AI en aan bijvoorbeeld de bedrijven en de systeembouwers die zich daarmee bezig houden. Dan nog moeten we scherp zijn op de democratische en rechtsstatelijke gevolgen, en op de legitimiteit. Maar het systeem zelf *kan* inhoudelijk perfect en volledig zijn. En hoewel die mogelijke/wellichte toestand nog ver weg in de tijd blijkt, gaat het hier om de vraag of dat de uiteindelijke toestand is waar we op denken of menen te koersen. We leven in de tussentijd van technologie, maar in deze eerste opvatting is dat dus de tijd tussen nu en uiteindelijke perfectie.

In het tweede geval ligt dat anders: het systeem is dan per definitie onvolledig, ziet per definitie dingen over het hoofd en maakt per definitie ongelukken onderweg. Dan is het niet verstandig om de sleutels over te geven aan de systeembouwers. En al helemaal niet aan bedrijven. Simpelweg, omdat we dan definitief het maken van politieke keuzes verplaatsen vanuit het systeem van democratische en rechtsstatelijke controle, naar de private belangen van systeembouwers en bedrijven. Google kan prima het roer overnemen op terreinen waar politieke verschillen geborgd zijn en een voor iedereen optimale verdeling te garanderen is: in a-politieke systemen dus. Maar als die systemen *niet* a-politiek zijn, dan betekent de 'outsourcing' van het

systeem feitelijk de privatisering van de politiek. En dat zou vanuit de democratische rechtsorde niet de bedoeling moeten zijn. In dat geval is de tussentijd waarin we leven er een waarin we niet toe bewegen naar uiteindelijk perfectie, maar waarin er per definitie, en dus altijd, imperfecties zullen zijn die bijsturing en overheidsinterventie vereisen. Het is dan een overgangsfase waarin we toewerken naar een wereld die weliswaar méér technologie en AI gedreven is, maar waarin de rol van de overheid en politiek *altijd* onverminderd nodig blijft om de technologie en de bijbehorende bedrijven te controleren en te sturen.

In beide scenario's is AI een disruptieve technologie met enorme mogelijkheden en enorme impact voor de inrichting van democratie, rechtsstaat en sturing. In beide gevallen is die beweging ook écht discontinu en *anders*, maar wel op een totaal andere manier. Met ook totaal andere gevolgen voor wat de overheid met technologie zou moeten doen. En met een heel andere betekenis van de tijd op weg daar naartoe, de *tussentijd*. De tijd waarin we nu leven. Als we onderweg zijn naar perfectie, dan is de tussentijd een te overbruggen periode, op weg naar een periode waarin technologie leidend kan en mag zijn. Als we onderweg zijn naar permanente imperfectie, dan zal technologie nooit werkelijk leidend *mogen* zijn en is het nu dus ook gevaarlijk en niet wenselijk om al te zeer in die richting te bewegen. Dan zal de overheid *zelf* moeten blijven werken aan een eigenstandige machtspositie tegenover de technologie-oligopolies. Dan zal de overheid zich *niet* moeten overleveren aan de belangen die daar gelden en die via de toepassingen van technologie worden verzilverd. Waar de tussentijd op weg naar perfectie vraagt om *meebewegen*, gaat het onderweg naar permanente imperfectie juist om *terugduwen*. Dezelfde tijd, een heel andere beweging. En dat maakt van deze vraag dus de kernvraag voor de overheid in de tussentijd: waar zijn we naartoe onderweg? Naar aanstaande digitale perfectie, of naar permanente digitale imperfectie?



# 6. Handelingsperspectief: strategieën voor het omgaan met AI in de tussentijd

## Strategisch omgaan met brede, diepe en onbedoelde gevolgen

AI is een ongekende technologie. Dat wil zeggen, we weten wat het *nu* zo ongeveer is en welke vervolgstappen er in de maak zijn. Maar waar die stappen toe leiden, en wat daarvan de gevolgen zijn, *kunnen we niet kennen*. De realistische manier van kijken naar de toekomst van AI is dat we niet weten waar het heen gaat. Het potentieel van de technologie is letterlijk ongekend en niet te voorzien. Dat is geen tekortkoming in het vooruitzien, maar de essentie van de beweging. Die onvoorspelbaarheid betekent dat we er *voorzichtig* mee om moeten gaan. Letterlijk *voor-zichtigheid*: goed voor je kijken terwijl je loopt, en onderweg zorgvuldig de volgende stappen plannen. Niet terughoudend of belemmerend, maar in het volle besef dat het *anders* kan uitpakken dan we denken én dat het majeure impact *zal* hebben op de kernprocessen en kernwaarden van het openbaar bestuur, de democratie en de rechtsstaat. Juist omdat we de toekomst van technologie niet kunnen voorzien moeten we er voor-zichtig mee omspringen.

De voor-zichtigheid is geboden, omdat technologie voor het eerst in onze geschiedenis de potentie heeft om *buiten* het klassieke model te treden, waarin menselijke politiek de instrumentele inzet van technologie controleert. De technologie vormt dan niet langer een invulling of aanvulling op menselijke politiek, maar treedt daar deels buiten en staat daar op delen ook boven. Daarmee wordt het democratische en bureaucratistische model van de afgelopen eeuw het object van vernieuwing en verandering, inclusief de instituties, ordeningen en regels die we hebben om de orde in dat model te bewaken.

Die potentie tot fundamentele verandering geldt op twee niveaus. Ten eerste zien we dat het belangrijk is om te onderkennen dat de gevolgen van de introductie van AI en digitale technologie vaak **breder zijn dan bedoeld/gedacht**. Veel technologie wordt strategisch, maar instrumenteel, ingezet om bepaalde vraagstukken aan te pakken en van nieuwe of betere oplossingen te voorzien. Technologie wordt dan ingezet binnen bestaande gedefinieerde problemen en binnen bestaande ideeën over beleid en uitvoering. Tegelijkertijd blijkt daarbij vaak dat de gevolgen breder uitwaaiëren, naar

andere beleidsdomeinen, of op andere maatschappelijke niveaus. Denk bijvoorbeeld aan de risicosignalering op basis van big data en AI. Die zorgt enerzijds voor een grotere pakkans van fraudeurs, maar maakt anderzijds dat bepaalde groepen systematisch worden uitgelicht en daar ten onrechte nadeel van ondervinden. Algoritmen hebben een zichzelf versterkend effect, of zorgen voor onbedoelde dynamiek, ver voorbij de strategisch beoogde inzet van het instrument. Bijvoorbeeld dat bepaalde groepen gebieden in de stad gaan mijden, bepaalde regelingen niet meer gebruiken, of bepaalde contacten voortaan geheimhouden, om uit de greep van de algoritmen te blijven. Let wel, dat is iets anders dan dat ze zich onthouden van criminele of illegale activiteiten. Ze zien af van rechten en ruimte die ze hebben, uit angst voor overmatige controle en proactieve en preventieve ingrepen. Voor veel inzet van technologie is die performatieve werking eerder regel dan uitzondering.

Ten tweede zien we dat gevolgen van technologie vaak een **dieper dan bedoeld** effect hebben; ze raken aan wortels en principes van het systeem, op een dieper niveau dan bij de inzet of uitrol ervan werd gedacht. Instrumentele inzet van data en AI kan bijvoorbeeld maken dat rechtsgelijkheid, rechtsbescherming of de rol van de politiek op de achtergrond raakt. De toepassing is dan weliswaar instrumenteel, maar de gevolgen zijn dat niet. En in die diepte-effecten zit ook weer gradatie, waarbij het lastig is om die te rangschikken naar ernst – omdat in principe elk onbedoeld effect op de principes van het systeem ernstig is. Die ernst is sowieso serieus. Niet vanuit een soort conservatief standpunt dat alle bestaande principes behouden moeten blijven, maar vanuit het idee dat veranderingen in principe alleen vanuit een expliciete en politieke route zouden moeten plaatsvinden. Principes verander je niet impliciet als onbedoelde bijvangst, maar in de volle openheid en transparantie van het politieke, democratische en rechtsstatelijke proces dat ervoor is ingericht.

Voor een deel zijn er ook diepte-effecten die relatief los staan van de instrumentele inzet van AI. Technologie heeft een **eigen inherente dynamiek**, die grote gevolgen heeft voor de democratie en voor de politiek. Technologie heeft bijvoorbeeld effect op de manier waarop de democratie werkt. We hebben gezien hoe sociale media leidt tot de vorming van bubbels en steeds fijnere micro-targeting van doelgroepen met gepersonaliseerde boodschappen mogelijk maakt. Dat heeft grote effecten op hoe verkiezingen uitpakken. Maar breder dan dat, ook op hoe de democratische orde en de samenleving worden beleefd. Sociale media is maatschappelijk transformatief, op een manier die moeilijk in beeld te brengen is, totdat gevolgen zich heel duidelijk manifesteren. Het analyseren van de gevolgen van digitale technologie is vaak als sterrenkijken. Wat je nu aan het firmament ziet zijn de gevolgen van wat lang(er) geleden is gebeurd. De gevolgen die we nu van

sociale media waarnemen, bijvoorbeeld in verkiezingen, het ontstaan van diepe bubbels en ‘rabbit-holes’, en het uiteenvallen van sociale cohesie, zijn gevolgen van wat jaren hiervoor gebeurde. Die gevolgen zijn enorm en niet bewust zo bedoeld. Het is geen gevolg van inzet van technologie door de overheid, maar het laat diepe sporen na in de samenleving waarop het overheidsbestuur en de democratie betrekking hebben. Het lijkt soms zelfs de vraag of we nog van een *samenleving* kunnen spreken. Of dat het in plaats daarvan meer gaat om een geografisch samenvallen van de plaats van leven, van verder niet met elkaar verbonden groepen of individuen. We leven samen, in hetzelfde gebied, maar verder is er amper samenhang of gevoelde lotsverbondenheid. De samenhang die we wél voelen is er dan één van onlinegemeenschappen en door gepersonaliseerde content gecreëerde gevoelens van gemeenschap met gelijkgestemden. Individuen leven in bubbels samen met de voor hen gepersonaliseerde content, die ze als gelijkgestemde anderen herkennen en aanvaarden. *Wat je ziet ben jezelf*, schreef Henri Beunders in 2000 al naar aanleiding van het toen nog spraakmakende programma *Big Brother*: digitale technologie heeft die dynamiek nog verder versterkt. En dat heeft onze samenleving – waarschijnlijk onbedoeld – enorm veranderd.

AI en nog verdere ontwikkelingen rond data en algoritmen kunnen – zullen! – dat proces van niet gevraagde diepe maatschappelijke gevolgen versterken en versnellen. Bijvoorbeeld in de fundamenteën van de democratie, het politieke en de werking van de rechtsstaat. Zo is ‘fake’ als bijvoeglijk naamwoord steeds lastiger houdbaar. Het impliceert dat het één echt is en het ander nep. Echt en nep komen in een virtuele en digitaal door AI gegenereerde wereld steeds meer over elkaar heen te liggen en lopen in elkaar over. Zo ontstaat een samenleving waarin verschillende eieren als in een omelet zijn toegevoegd. Die verschillende eieren gaan er wel als separate elementen in, maar zijn vervolgens niet meer in de omelet te onderscheiden – laat staan te scheiden. Dat heeft niet alleen diepe effecten in de principes en fundamenteën, maar deels ook *ondermijnende* effecten voor de democratie en rechtsstaat. Ook dat is een categorie gevolgen die we serieus moeten nemen. Helemaal vanwege de in dit essay ook besproken partijen die eigenaar zijn van veel van de data en algoritmen: private bedrijven, autocratische regimes, en dikwijls ook allianties tussen die twee. Waarbij geldt dat de aard van de techniek het voor die partijen en hun toepassingen mogelijk maakt om zich aan de rechtsorde en de statelijke structuren en regulerende instituties te onttrekken, wat hun ondermijnende werking lastig te controleren en nog veel lastiger te beheersen maakt.

Zo stralen de gevolgen van AI en technologie breder en dieper uit dan we vooraf kunnen voorzien. Dat vereist oplettendheid, waakzaamheid en zorgvuldigheid

bij het verder introduceren en/of toelaten van AI in het openbaar bestuur. De publieke waarde van AI is nooit alleen maar positief: de winst voor de één is vaak een verlies voor een ander. Voortgang op het ene domein, kan nieuwe en diepere problemen op andere terreinen betekenen. De toevoeging van AI en technologie *verandert de verdeling van winst en verlies*. Dat is een heel andere benadering dan de protagonisten van technologie vaak hanteren. Die stellen dat technologie de wereld verbetert, voor iedereen. Een dergelijke voor iedereen positieve uitkomst bestaat in het openbaar bestuur en het publieke domein zelden. Paradoxaal genoeg vormen AI en technologie een existentiële bedreiging voor een aantal fenomenen die wél de conditie van iedereen verbeteren, zoals een goed werkende democratie, de rechtsstaat, een goede publieke dienst en een verbonden en betrokken samenleving. Er is niets principieels tegen het veranderen van een verdeling, maar dat moet wel expliciet gebeuren. Niet impliciet en zeker niet via de band van de introductie van “simpelweg” nieuwe en neutraal veronderstelde instrumenten.

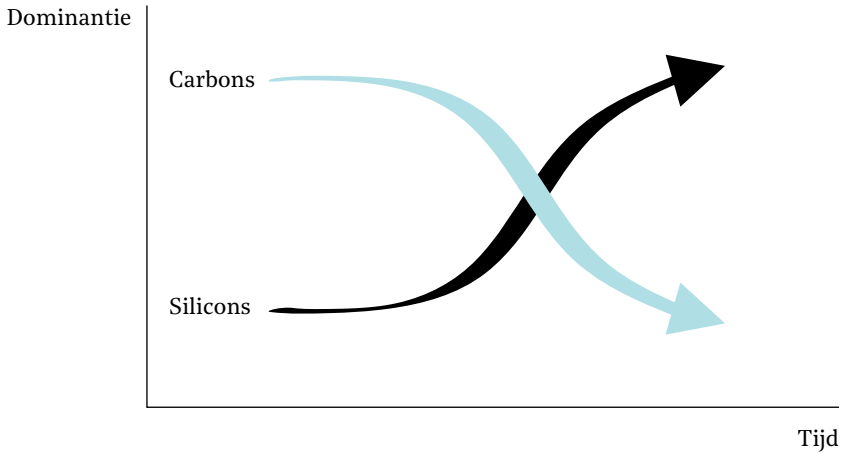
Dit essay biedt langs twee routes ‘checklists’ voor zorgvuldige en gecontroleerde introductie van AI en technologie in het publieke domein, die passen in een voor-zichtige strategie van de overheid. De overheid staat niet met lege handen. Het gaat vooral om het bewustzijn van de hier genoemde bredere, diepere en onbedoelde gevolgen – niet als uitzondering maar als regel. *The purpose of the system is what it does*, biedt een handzaam aangrijpingspunt voor sturing. Niet teveel filosoferen over wat het zou kunnen zijn, maar vooral scherp, oplettend en realistisch kijken naar wat er gebeurt en wat technologie in de praktijk doet. En dan bijstellen, aanpassen, doorzetten of indien nodig terugtrekken. Permanent onderhoud dus en actief uitschakelen van de ‘bedoeling-bias’ door middel van de checklists in dit essay. Een politieke vraag daarbij is of de checklist daarbij wordt gehanteerd als *ja, mits* of als *nee, tenzij*. Allebei is mogelijk, maar legt een andere bewijslast voor de introductie van AI en digitale technologie. Hoe hoger de bewijslast, hoe minder open en zoekend innovatie kan zijn. Dat verkleint de kans op ongelukken, maar het verkleint ook de kans op verrassende positieve gevolgen. Voor de weging van die criteria is geen objectieve maatstaf te formuleren. Het is een strategische vraag die beleidsmakers en ontwerpers van arrangementen zich moeten stellen: welk criterium prevaleert hier?

## Strategie voor de tussentijd

Tenslotte, maar volgens mij als eerste en belangrijkste, is er nog de onzekerheid over de uiteindelijke belofte van AI en technologie: is perfectie – in de zin van acceptabele hanteerbaarheid van alle (!) maatschappelijke complexiteit – mogelijk, of gaat het uiteindelijk om permanente imperfectie?

Kunnen we volledig zelfrijdend vervoer bereiken, of zijn er altijd mensen nodig om in bepaalde situaties het stuur over te nemen? Dat is een belangrijke vraag, omdat het antwoord op die vraag het verschil betekent tussen enerzijds het langzaam *maar zeker* uitfasen van de carbons uit het systeem, en anderzijds het bouwen van een hybride systeem met daarin de mens als ultieme borger van waarden en belangen. In de volledig zelfrijdende auto (niveau 5) is de mens als relevante bestuurder uitgefaseerd, in het tot op één na hoogste niveau (4) van zelfrijdende technologie is de mens nog steeds de bepalende én verantwoordelijke figuur in de zo goed als geheel zelfrijdende auto. *Als* de techniek perfect kan zijn, *dan* is er weinig reden voor de mens om zijn prominente positie te handhaven; die is dan eerder een bron van fouten en verstoring dan van toegevoegde waarde. Dan is het mogelijk en misschien zelfs wenselijk om de rol van de mens steeds meer te vervangen door technologie. Als dat niet zo is, dan *moet* de mens uiteindelijk altijd in staat zijn en blijven om in te grijpen en te overzien en te begrijpen wat het systeem doet.

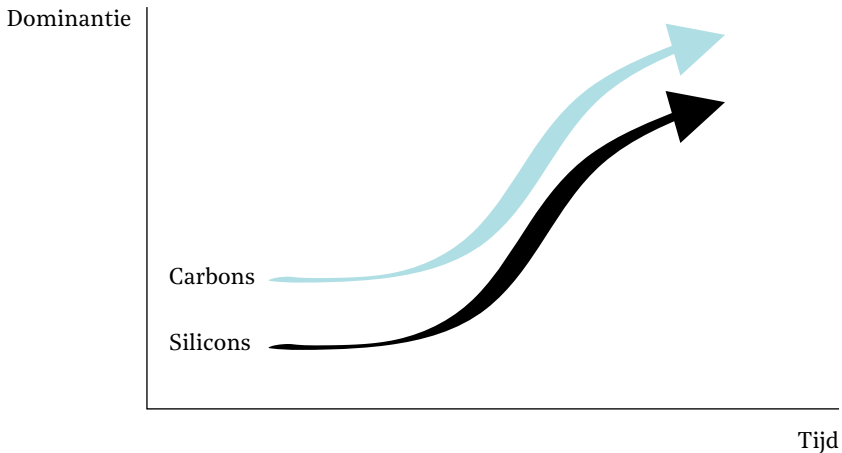
Daarbij hoort het besef dat veel van de technologie waarover we het hebben *niet* in handen is van publieke diensten, maar van private ondernemingen. En in het geval van China en verschillende andere landen in de handen van semi-staatsbedrijven, die de facto onderdeel zijn van het autocratische regime. Omdat de voor zelfsturing benodigde technologie zich ontwikkelt via data-accumulatie en padafhankelijkheid, en het oligopolie van een selecte groep marktpartijen dat daar inherent bij hoort, gaat het dan per definitie om macht die verschuift van de overheid naar een selecte groep bedrijven. Hier gaat het dus niet alleen om transfer van macht van carbon naar silicon, maar ook om overdracht van macht van de overheid naar de markt. Onderstaande figuur 2 illustreert die beweging, volgens de x-curve van transitie zoals Derk Loorbach en DRIFT die hanteren. De overheid bouwt in dit scenario op steeds meer terreinen zijn eigen rol af, en draagt die over aan technologie die de publieke taken op zich neemt en daarin meer publieke waarde verzorgt. De overheid bouwt af en geeft een belangrijk deel van de eigen rol en positie over aan de markt, waar de technologie beschikbaar is. Hier geeft een bestuursorgaan bijvoorbeeld de eigen taken uit handen aan Google of Meta, met het verzoek om de benodigde publieke waarde beter en anders te leveren. De tussentijd is dan de fase waarin de overheid langzaam maar zeker meer taken overdraagt aan technologie, op weg naar een technologie die het publieke domein en publieke waarden zelf stuurt en zelf realiseert – met de mens en de overheid als gebruiker, niet als bestuurder.



*Figuur 2: het x-curve scenario voor digitalisering: afbouw van carbon, opbouw van silicon.*

Maar wat als dat niveau van zelfsturing en zelfrealisatie van publieke waarde door technologie *niet* of *niet volledig* mogelijk is? Wat als we niet toewerken naar niveau 5 zelfrijden, maar niveau 4 het maximaal haalbare is? Dat betekent dat technologie nooit de perfectie kan leveren die het overgeven van het stuur vereist en *permanent imperfect* zal zijn. In dat geval is een afbouw langs een x-curve niet de gewenste strategie, maar is juist een dubbele s-curve nodig. Zie daarvoor onderstaande figuur 3, waarin de introductie van AI en digitale technologie weliswaar grote slagen maakt, máár daarbij altijd menselijke controle, sturing en menselijke interventie nodig en mogelijk blijft. In dat geval faseert de mens niet uit, maar groeit deze noodgedwongen mee. *Als* de techniek niet perfect is maar permanent imperfect – steeds op een hoger niveau, maar met daarbij ook op een hoger niveau imperfectie -, *dan* moet de mens er steeds op dat hogere niveau nog boven weten te hangen. Hier moet de mens in het bestuur de technologie en de markt dus te boven en vooruit blijven. Dat kan door de eigen ontwikkeling te versnellen, maar ook door de ontwikkeling en introductie van technologie in de markt gericht te vertragen. Dat heeft hele grote gevolgen, want het betekent dat de overheid in staat moet blijven om te controleren en snappen wat technologie en marktpartijen doen, kunnen en waar de beperkingen liggen. De overheid faseert niet uit, maar moet steeds een stap voor blijven. Niet perse in de uitvoering, maar wel in de kennis en het inzicht van wat er kan en gebeurt. Dus wellicht wél Google en Meta uitnodigen om actief te zijn, maar altijd met zelf het finale inzicht, de kennis en de zeggenschap. Waarbij altijd geldt dat zeggenschap zonder inzicht en kennis onzinnig is. Dat is dus ook meer dan een simpele intentieverklaring of een machtswoord. Wie technologie zinvol wil sturen kan niet anders dan zich erin verdiepen en radicaal investeren in het begrip ervan. Dat is niet alleen zeggen dat je ervan bent, maar betekent het

daadwerkelijk verdiepen en weten wat je doet. En dat kost in de wereld van technologie altijd héél veel geld en tijd: het vereist diepe zakken en onwrikbare wil, precies datgene dat in een politieke democratie niet vanzelfsprekend voorhanden is. Er zijn concurrerende prioriteiten en politieke stromingen. Dat maakt deze strategie voor overheden per definitie lastig. Niet onmogelijk, wel inspannend.



*Figuur 3: het dubbele s-curve scenario voor digitalisering: opbouw van silicon, maar carbon blijft in control en dus ook 'on top'.*

De x-curve en de dubbele s-curve zijn twee expliciete strategieën voor de overheid voor het omgaan met technologie, waarbij er sprake is van een bewust gekozen verhouding tussen het aandeel silicon en het aandeel carbon in het stelsel. Ze zijn allebei vormen van *voor-zichtigheid*, maar wel met een heel andere inzet en uitgangspunt. Zonder een dergelijke keuze ontstaat eerder een 'flatline' van het bestaande bestuur. Dan groeien de mogelijkheden van technologie boven de menselijke mogelijkheden tot controle en sturing uit, maar zonder idee over toekomstige perfectie of imperfectie. Dan laten we het over aan het toeval, in de wetenschap dat de markt zelden tot een oplossing met rechtvaardige verdeling van publieke waarde in de samenleving komt. In het gelukkige geval van perfectie is dat niet zo erg, in het ongelukkige geval van permanente imperfectie betekent dit dat niemand meer stuurt, controleert, borgt of anderszins democratische en rechtsstatelijke waarden bewaakt. In dat geval is het openbaar bestuur de facto 'dood'; het is er wel, maar kan zijn meest fundamentele en basale functie niet meer vervullen.

Bij al deze figuren geldt dat zowel marktpartijen als de overheid zelf in theorie de vergevorderde technologie en AI zou kunnen ontwikkelen. Maar de praktijk zal waarschijnlijk anders zijn. Door de enorme kosten voor de ontwikkeling

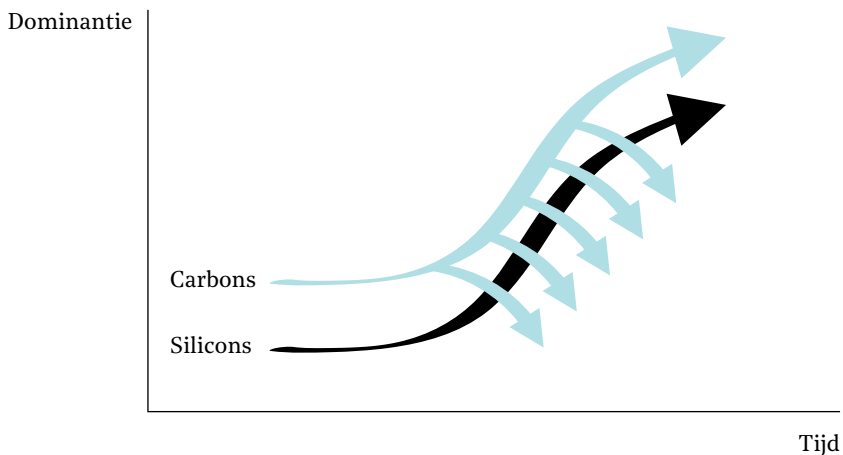
van de hoogwaardige programmatuur én de daarvoor benodigde enorme hoeveelheid data en gebruik, is concentratie bij een klein aantal marktpartijen (wellicht wel semi-staatsbedrijven, zoals we in bijvoorbeeld China zien) waarschijnlijker. Dit betekent dat de vraag of technologie perfectie zal bereiken eigenlijk gaat om de vraag of *bedrijven* die perfectie zullen kunnen bereiken. En dus ook om de vraag of die marktpartijen in de x-curve de overheid ‘vervangen’, of ze in de dubbele s-curve door de overheid gecontroleerd blijven, of dat ze de overheid in een ongecontroleerd en grillig proces verdringen. Nogmaals, dit is geen inherent gevolg van technologie, maar meer een logisch gevolg van de manier waarop het eigenaarschap van technologie, AI en data zich ontwikkelt. Daarin lopen bedrijven simpelweg voorop én steeds verder weg bij de overheid. Ze beheren de technologie, de data én de infrastructuur waarover die zich beweegt. Ze genereren daarmee kapitaal dat de investeringsmogelijkheden van overheden te boven gaat. In de praktijk beweegt de stroom zich dus in de richting van private bedrijven, niet in die van overheden. Die stroom is geen natuurwet. Hij is te keren, maar dat vereist gerichte en gedegen inspanning van de overheid zelf.

Strategie in de tussentijd vereist dus van overheden dat ze expliciteren in welk type curve ze denken te verkeren: waar werken ze naartoe en welke verwachting over technologie ligt daar onder? Dat hoeft geen generieke afweging of keuze te zijn, de strategie kan per onderwerp verschillen. De Tesla is op snelwegen zo goed als 100% zelfrijdend, maar op rotondes en in druk stedelijk gebied zeker niet. Op de snelweg laat ik het stuur los, in de stad rijd ik zelf. Overheden kunnen per onderwerp het gesprek voeren over de mate van perfectie die ze verwachten en daar hun strategie op aanpassen. Van daaruit kunnen ze verschillende kanten op. Ze kunnen op bepaalde dossiers het voortouw nemen in het actief – alleen of met marktpartijen – ontwikkelen en toepassen van technologie. De overheid kan vooroplopen en het pad aanvoeren, binnen de waarborgen die we daarvoor hebben geschetst. De overheid kan op andere dossiers ook actief terugduwen en de ontwikkelingen in de markt rondom bepaalde toepassingen of domeinen afsluiten. Niet alles wat de markt wil hoeft. Het is juist de taak van de overheid om vanuit de eigen strategische positionering kritisch te zijn in waar de toepassingen van technologie gewenst zijn én waar niet. De overheid kan daar assertief in zijn. Dat het elders wellicht wél gebeurt is geen valide argument. Het gaat om de bescherming van de eigen populatie en de eigen politieke en democratische gemeenschap.

Een derde strategie is zodoende wat in de medische wetenschappen *watchful waiting* wordt genoemd (zie figuur 4). Hierbij kiest de arts (of in dit geval de overheid) ervoor om actief niets te doen, aan te zien hoe het zich ontwikkelt en daar vervolgens op te acteren. Het is mijn vermoeden dat dit voor veel



inzet van technologie en AI een goede strategie is. Dus geen generieke x of s-curve, maar vanuit een s-curve zoals in figuur 4 gericht onderzoeken in welke domeinen of voor welke taken sporen van perfectie optreden en waar afbouw (en dus een specifieke x-curve) mogelijk is. Het is voor veel technologie niet goed te voorspellen naar welk eindstadium die toewerkt en in welke tussentijd we verkeren. Dat betekent dat het niet verstandig is om al finale keuzes te maken en al helemaal niet om maar gewoon de gok te nemen. Watchful waiting, het actief zoeken naar ontwikkelingen en signalen, is dan een goede strategie. Door gericht te kijken naar wat er gebeurt en wat dat voor gevolgen heeft, en daar op te reageren. Terughoudendheid dus ook, maar niet vanuit passiviteit. Soms, zo geldt ook voor de toepassing van technologie, is niets doen heel hard werken. Om goed *watchful waiting* te kunnen doen, moet de overheid weten en snappen wat in domeinen de stand van de technologie is. En alleen dat weten vereist al enorme investeringen en – zoals hiervoor al aangeduid – lange adem en continuïteit. De enorme kosten wegen wat mij betreft echter goed op tegen de baten. Het voorkomt het ‘per ongeluk’ verliezen van controle over kritieke verworvenheden als democratie, rechtsstaat en zeggenschap over het publieke domein. Het stelt bovendien in staat om waar het wél kan maximaal de vruchten van digitalisering en AI te plukken. De overheid kan gerichte taken afbouwen en overdragen aan technologie, maar altijd vanuit een expliciet eigen besef van wat daarvan de kansen en beperkingen zijn.



*Figuur 4: het watchful waiting scenario voor digitalisering: carbon 'on top' en in control, met selectieve afbouw ten gunste van opbouw van silicons.*

We begonnen dit essay met de Tesla en de niet-willekeurige maar ook niet-verklaarbare noodstop op de snelweg. Net als met de zelfrijdende auto is AI een technologie in de tussentijd: onderweg naar een volgend normaal, waarin het een belangrijk deel van ons leven en ons openbaar bestuur zal bepalen. Maar de tussentijd is geen passieve fase, waarin we kunnen afwachten. De keuzes die we nu maken zullen bepalend blijken te zijn voor de toekomstige verhouding tussen overheid en marktpartijen, tussen siliconen en carbons, en voor de toekomstige verdelingen tussen maatschappelijke groepen en hun belangen. En hoewel we niet kunnen voorspellen of de toekomst van technologie ooit *perfect* zal zijn of niet, zijn de keuzes ten aanzien van technologie en de bedrijven die technologie aanbieden, ontwikkelen en beheren voor beide scenario's heel anders. In een scenario van toekomstige perfectie is er niet zoveel mis met het gradueel overlaten van maatschappelijke en politieke ordening aan tech-partners, en het overgeven van het stuur aan digitale technologie en AI. In een scenario van permanente imperfectie is dat zeer problematisch en is het nodig om als overheid vooruit te blijven en macht ten koste van technologie (bedrijven) te behouden. In het ene geval is de transitie een X, in het andere geval een dubbele S. Dat zijn niet zomaar heel andere letters, maar heel andere manieren van doen en denken, en van sturen en investeren. Dat geldt niet alleen als meta-strategie voor de lange termijn, maar laat zich voelen in allerlei dagelijkse keuzes over technologie. Die vandaag op de agenda (zouden moeten) staan. Over de mate van toegang en controle die technologie (bedrijven) krijgen over delen van het openbaar bestuur, over het wel of niet plaatsen van sensoren, het wel of niet in gebruik nemen van een voorspellend algoritme, het wel of niet experimenteren met Smart City concepten of digital twins.

In dit essay heb ik niet geprobeerd om de toekomst van AI te voorspellen, maar om de consequenties van de mogelijke toekomst te schetsen. Welke strategie overheden van daaruit kiezen is aan hen, idealiter als uitkomst van politieke deliberatie en democratische besluitvorming. Dat zijn debatten die we meer, vaker, fundamentele en strategischer moeten voeren. De Tesla corrigeert zijn remfouten bij een volgende update. Het is tijd voor ons als carbons om kritisch te kijken naar ons eigen presteren ten aanzien van digitale technologie en daar – net als Tesla – van te leren en onszelf waar nodig te corrigeren. Het pad naar de toekomst is te ongewis om aan de siliconen over te laten. *Tussentijd vereist voor-zichtige strategie van carbons.*

# Handelingsperspectief

## 1. Bewaak democratische en rechtsstatelijke kwaliteit van AI-toepassingen

- AI is wat het doet, niet wat de bedoeling ervan was.
- Vier vragen om democratische en rechtsstatelijke kwaliteit van AI-toepassing te toetsen:
  - Is het politieke van de toepassing onderwerp van politieke deliberatie geweest?
  - Zijn rechtszekerheid en rechtsbescherming van subjecten geborgd?
  - Is rechtsgelijkheid geborgd, in het ontwerp én in de uitkomsten?
  - Is de legitimiteit van een toepassing op orde voor alle (!) vier de vormen van legitimiteit (input, throughput, output, outcome)?
- Bij twijfel: zet stop, herontwerp en corrigeer, en start desgewenst (snel) weer op.
- Niet sneller dan mogelijk, maar ook niet langzamer dan nodig: bouw democratische structuren voor snelle bijstelling.

## 2. Stuur de inzet van AI-toepassing bij de realisatie van publieke waarde

- Doorloop de publieke waarde cyclus linksom: eerst de brede autoriserende omgeving, daarna pas de organisatorische omgeving.
- Wees kritisch op rechtsdraaiende voorstellen: op AI die bestaande organisatiemodus sublimeert en de belangen van een selecte autoriserende omgeving behartigt.
- Het effect van AI op publieke waarde is altijd dieper en breder dan bedoeld: stel permanent bij.

## 3. Investeer in strategische capaciteit om de onvermijdelijke disruptie van AI te sturen

- Expliciteer het verwachte eindstadium van AI per vraagstuk: uiteindelijke perfectie, of permanente imperfectie. Varieer per vraagstuk.
- Kies een gewenst scenario voor verhouding siliconen en carbons: x-curve, dubbele s-curve, of watchful waiting.
- Géén keuze is een keuze voor verlies van positie van carbons en van de overheid.
- De keuze voor de dubbele s-curve en/of watchful waiting vereist bovenliggende eigen kennis van AI en digitale technologie: dus, radicaal investeren in eigen kennis van AI en allianties aangaan met trusted publieke partners, zoals universiteiten en kennisinstellingen.

# Referenties

- Berg, M. van den (2023). *De beleidsbubbel. En hoe we die liefdevol leeg laten lopen*. Deventer: Verhaal met impact.
- Berg, C.F. van den, M. van der Steen, L.G. Tummers (2021). *Openbaar bestuur: beleid, organisatie, politiek*. Deventer: Wolters Kluwer.
- Frissen. P.H.A. (2023). *De integrale staat. Kritiek van de samenhang*. Den Haag: Boom.
- Huiting, M., W. Hendriks, H. den Uijl, A. Frankowski, M. Schulz, P. Frissen, M. van der Steen (2023). *De (on)begrensde menselijke maat*. Den Haag: NSOB.
- Moore, M.H. (1995). *Creating Public Value: Strategic Management in Government*. Cambridge: Harvard University Press.
- Kuipers, G., G.J. Geertjes & M. van der Steen (2023). *Wie wordt gehoord? Dynamiek in het landschap van macht en tegenmacht*. Den Haag: NSOB.
- Maes, R., E. de Vries, W. Bronsgeest & M. Wesseling (2017). *De informatieprofessional 3.0*. Meppel: Boom.
- Poli, R. (2019). *Handbook of Anticipation: Theoretical and Applied Aspects of the Use of Future in Decision Making*. Berlin: Springer.
- Schuilenburg, M. *Making surveillance public. Why you should be more woke about AI and algorithms*. Den Haag: Eleven.
- Schuilenburg, M. & R. Peeters eds. (2020). *The Algorithmic society. Technology, Power, and Knowledge*. London: Routledge.
- Schulz, M., M.J.W. van Twist, M. van der Steen, C. van der Kaaij (2023). *Intelligent bestuur verbeeld(t). Modelleren van openbaar bestuur*. Den Haag: NSOB.
- Snowden, D. (2002). *Complex Acts of Knowing: Paradox and Descriptive Self-Awareness*. Journal of Knowledge Management 6, no. 2: 100–111.
- Steen, M. van der (2016). *Tijdig bestuur. Strategisch omgaan met voorspelbare verrassingen*. Den Haag: NSOB.

- Steen, M. van der, G. Boogaard, P. Jansen, J. Westerweel, B. Koopmans (2021). *Legitimititeit gezocht: dilemma's van democratische afstand en nabijheid in het RES-proces*. Den Haag: NSOB.
- Steen, M. van der, J. Burger, W. Burger (eds.) (2018), *The governance of transitions; the transition of governance*, Stellenbosch University Press.
- Steen, M. van der, M. van Delden, M. van 't Wout, J. Oudega (2023). *Een waterschap voor het waterbeheer van de toekomst: strategie als gedeelde betekenisgeving door de Nederlandse Waterschappen*. NSOB: Den Haag.
- Steen, M. van der, A. Frankowski, L. Schröer, J. Scherpenisse, N. Chin-A-Fat (2022). *Omgaan met Signalen: naar een handelingsperspectief voor overheden*. Den Haag: NSOB.
- Steen, M. van der, J. Scherpenisse & M.J.W. van Twist (2015). *Sedimentatie in sturing. Systeem brengen in netwerkend werken door meervoudig organiseren*. Den Haag: NSOB.
- Steen, M. van der, J. Schram & J. Scherpenisse (2018). *Het waarderen van wonen. Een reflectie op het instrument van maatschappelijke visitaties van woningcorporaties*. Den Haag: NSOB.
- Steen, M. van der & M.J.W. van Twist (2020). *Foresight and Anticipation: strategies for learning ahead of time*. In: Greve, C. (ed.), *The Oxford Handbook of Policy & Time*, Oxford University Press, Oxford.
- Teisman, G., A. van Buuren & L. Gerrits, eds. (2009). *Managing Complex Governance Systems*. New York: Taylor & Francis.
- Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks: Sage.
- Weick, K. E., & K. M. Sutcliffe (2007). *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. San Francisco: Jossey-Bass.

# Eerder publiceerde de NSOB

2024

## Onderweg ontwerpen

Een stelsel voor onderwijs aan nieuwkomers ontwerpen

*Annemarie van der Wilt, Martijn van der Steen, Christiaan van der Kaaij,  
Eva Kloet, Laura Schröer, Bart van Kessel*

## Audit, maar dan anders

Een verkenning naar hoe het repertoire van auditors zich laat verbreden en verdiepen

*Jorren Scherpenisse, Mark van Twist, Robin Hill*

## Kostbare geneesmiddelen

De VIG tussen leden, politiek en maatschappij

*Andrea Frankowski, Paul Frissen, Tom Lubbers, Martin Schulz, Mark van Twist*

## Gedeelde sturing

Regionaal samenwerken voor een betere aansluiting tussen mbo en arbeidsmarkt

*Annemarie van der Wilt, Tijs van de Vijver, Laura Schröer, Jessie Samwel,  
Wiljan Hendriks, Nancy Chin-A-Fat*

## Werk maken van erkenning

Over contact als toegangspas

*Guido Rijnja*

## Leren van onderzoek naar ongevallen

Hoe onderzoekscommissies een narratief construeren en wat dat betekent voor leren

*Hans de Bruijn*

## Signalen managen

Over de omgang met signalen bij het ministerie van SZW

*Andrea Frankowski, Laura Schröer, Annemarie van der Wilt, Wiljan Hendriks,  
Robin Hill, Myrthe van Delden, Martijn van der Steen*

### **Over intenties en interferenties**

Een lerende evaluatie van de schadeafhandeling rond Kanaal Almelo-De Haandrik

*Myrthe van Delden, Georgina Kuipers, Mark van Twist, Marise van 't Wout, Robin Hill*

### **Bedreiging en bescherming**

Tussenevaluatie van de netwerkaanpak van het Netwerk Weerbaar Bestuur

*Georgina Kuipers, Gert Jan Geertjes, Jessie Samwel, Christiaan van der Kaaij, Rogier van der Wal, Willeke Slingerland en Martijn van der Steen*

### **Kwesties op de kaart**

Strategische ontwikkelrichtingen voor de netwerkaanpak van droogte in de Achterhoek en de Liemers

*Myrthe van Delden, Christiaan van der Kaaij, Mark van Twist, Martin Schulz*

### **Vormgeven aan schadebeleid**

De zoektocht van de Belastingdienst naar vertrouwenwekkend schadebeleid voor de Fraude Signalering Voorziening

*Georgina Kuipers, Marije Huiting, Martijn van der Steen, Myrthe van Delden*

### **Samenwerken in een brede coalitie**

Acht lessen over de totstandkoming van de Werkagenda mbo en het Stagepact

*Wiljan Hendriks, Andrea Frankowski, Robin Hill, Jessie Samwel*

## **2023**

### **What got us here, won't get us there**

Ver snellen van de transitie naar Circulair Zuid-Holland

*Martin Schulz, Marije Huiting, Martijn van der Steen, Robin Hill*

### **Vormgeven aan interactie**

Hoe overheid en maatschappelijke initiatiefnemers in de praktijk tot samenwerking komen bij leefomgevingsvraagstukken

*Jorgen Schram, Eva Kunseler, Pia Nabielek, Martin Schulz, Mark van Twist*

### **Sport in beweging**

*Wiljan Hendriks, Georgina Kuipers, Martijn van der Steen,*

*Rachel-Roxelane Speelman, Anika Scholten*

### **De gemeentesecretaris buiten de deur**

Vragen bij een antwoord

Mark van Twist, Marije Huiting, Martin Schulz

### **De (on)begrensde menselijke maat**

*Marije Huiting, Wiljan Hendriks, Henk den Uijl, Andrea Frankowski,*

*Martin Schulz, Paul Frissen, Martijn van der Steen*

### **Toekomst geven aan tussenland**

Samenwerking bij het (verder) vitaliseren van vakantieparken op de Veluwe

*Laura Schröer, Martin Schulz, Mark van Twist*

### **Wie wordt gehoord?**

Dynamiek in het landschap van macht en tegenmacht

*Georgina Kuipers, Gert Jan Geertjes, Martijn van der Steen*

### **Intelligent bestuur verbeeld(t)**

Modell(er)en van openbaar bestuur

*Martin Schulz, Mark van Twist, Martijn van der Steen, Christiaan van der Kaaij*

### **Afwachten en anticiperen**

Hoe de gemeente Amsterdam stuurt op ontwikkelingen in het digitale publieke domein

*Georgina Kuipers, Christiaan van der Kaaij, Henk den Uijl, Valerie Frissen,*

*Jorgen Schram*

### **Een waterschap voor het waterbeheer van de toekomst**

Strategie als gedeelde betekenisgeving door de Nederlandse waterschappen

*Martijn van der Steen, Myrthe van Delden, Marise van 't Wout, Johan Oudega*

## **2022**

### **Ambtelijke ruimte**

Kan de overheid de grote maatschappelijke vraagstukken aan?

*Henk den Uijl, Paul Frissen, Martin Schulz, Johan Oudega*

### **Het verbinden van botsende institutionele logica's**

Een typologie van strategieën

*José Nederhand, Martijn van der Steen, Mark van Twist*



### **Omgaan met onbegrepen gedrag**

Dilemma's in de samenwerking en informatie-uitwisseling rondom personen met verward gedrag

*Christiaan van der Kaaij, Ninske Repko, Jorren Scherpenisse, Laura Schröer, Martin Schulz, Mark van Twist*

### **Governance voor transities**

Naar coping strategieën voor de omgang met governance dilemma's bij de transitie naar een circulaire economie

*Johan Oudega, Martin Schulz, Mark van Twist*

### **Omgaan met signalen**

Naar een handelingsperspectief voor overheden

*Martijn van der Steen, Andrea Frankowski, Laura Schröer, Jorren Scherpenisse, Nancy Chin-A-Fat*

### **Overleg in beweging**

30 jaar overleg tussen overheid en samenleving in de fysieke leefomgeving

*Martijn van der Steen, Marije Huiting, Myrthe van Delden, Christiaan van der Kaaij, Marise van 't Wout, Mark van Twist*

### **Verantwoorden over vertrouwenswerk**

Een kwestie van tellen én vertellen

*Georgina Kuipers, Myrthe van Delden, Mark van Twist*

### **Maatwerk als standaard**

Over dilemma's en dynamieken bij het streven naar 'maatwerk' in de uitvoeringspraktijk

*Wiljan Hendriks, Marije Huiting, Caspar van den Berg, Mark van Twist*

### **Stadsmariniers in de spiegel**

Reflecties op de veranderende rol van stadsmariniers in Rotterdam en Antwerpen

*Jorren Scherpenisse, Mark van Twist, Sebastian Wijnands, Rike Hendrikse*

### **De stroom zoeken**

De energietransitie bezien vanuit ecosysteem perspectief

*Martijn van der Steen, Maarten Otto, Georgina Kuipers, Sander Oosterloo, Petra Ophoff, Alexander Woestenburg*

***Publicaties uit eerdere jaargangen zijn te vinden op [www.nsob.nl](http://www.nsob.nl)***

**COLOFON**

© 2024

ISBN NUMMER

978-90-834135-9-4

ONTWERP

Ontwerpwerk, Den Haag

---

# Nederlandse School voor Openbaar Bestuur

[www.nsob.nl](http://www.nsob.nl)

## Over NSOB

De Nederlandse School voor Openbaar Bestuur (NSOB) is een hoogwaardig onderwijs- en onderzoeksinstituut in het hart van de hofstad. Binnen dit instituut worden wetenschappelijke kennis over en praktijkervaring met het Nederlandse openbaar bestuur met elkaar verbonden. De NSOB organiseert opleidingen, publieke debatten en verzorgt wetenschappelijke en professionele publicaties.