



‘Aan data zit een politieke en ethische component.’

‘Data en algoritmes bieden veel mogelijkheden, maar ze zijn niet neutraal. Wat je onderzoekt, welke algoritmes je bouwt, welke data je gebruikt en waarvoor, dat zijn politieke en ethische keuzes. Beleidsmakers zijn zich daar onvoldoende van bewust. Ook voor lokale politici is basiskennis van datageletterdheid absoluut noodzakelijk.’ Dat zegt de Nederlandse data-expert **Mirko Tobias Schäfer**.

Mirko Tobias Schäfer:
‘We moeten nu het
bewustzijn ontwikkelen
dat open standaarden en
transparante algoritmes
noodzakelijk zijn voor
een democratie.’

Mirko Tobias Schäfer is hoofddocent nieuwe media en digitale cultuur aan de Universiteit Utrecht en hoofd van de Utrecht Data School. Hij onderzoekt de maatschappelijke impact van datapraktijken. ‘Van data en algoritmes wordt vaak aangenomen dat ze neutraal zijn, maar dat is lang niet altijd zo. Data zouden de nieuwe olie zijn, maar die vergelijking gaat niet op. Data zitten niet in de grond, iemand heeft beslist om bepaalde data te verzamelen en te bewaren en andere niet. Er zit een politieke en ethische component aan vast. Big data raken ons begrip van cultuur, van burgerschap, van democratie. Daarom is het zo belangrijk dat ook de geesteswetenschappen zich hiermee bezighouden en duiding geven bij het implementeren van deze praktijken, nadenken over wat wenselijk is en wat niet. Ik ben een mediawetenschapper, geen digitaal expert.’

‘Een dataproject opstarten is een politieke beslissing; wat is het probleem en hoe kan een dataproject ons helpen om het op te lossen? Doen we het of doen we het niet? Hoe doen we het? Politici moeten ook de resultaten interpreteren, ze moeten dus weten wat er achter de data zit.’

Zijn data per definitie niet neutraal?

‘Nee. Er zijn algoritmes die geen impact hebben op mensen, die bijvoorbeeld heel nauwkeurig voorspellen welke watermolen wanneer niet meer zal werken of aan vervanging toe is. Maar veel data zijn niet neutraal. Het is goed dat er nu veel aandacht is voor die valkuilen. Auteurs als Cathy O’Neil of Virginia Eubanks tonen overtuigend aan dat er problemen zijn met algoritmes en data, dat ze racistisch en bevooroordeeld kunnen zijn, dat ze ongewenste sociale opvattingen kunnen versterken. Maar dat negatieve beeld mag ons niet doen vergeten dat algoritmes en data dingen mogelijk maken die vroeger niet konden. Ze geven de kans te reageren op nieuwe uitdagingen, ze openen nieuwe mogelijkheden om problemen op te lossen, ze leiden tot nieuwe, multidisciplinaire vormen van samenwerken. We moeten wel beseffen dat ze politieke verantwoordelijkheid vragen, dat we begrip moeten ontwikkelen over hoe datapraktijken of algoritmes werken, hoe een dataset in elkaar zit.’

Kunt u enkele voorbeelden geven van hoe data bevooroordeeld kunnen zijn?

‘De leefbaarheidsbarometer van de Nederlandse overheid brengt de leefbaarheidssituatie per wijk in beeld. De wijk in Rotterdam waar ik vele jaren



heel graag heb gewoond, haalde een lage score. Nu blijkt dat het aantal allochtonen in een wijk een van de factoren is die het algoritme meeneemt in het bepalen van de leefbaarheid. Ook in Nederland maakt de politie gebruik van een criminaliteitsanticipatiesysteem. Het resultaat is een kaart waar per gebied wordt aangegeven hoe groot de kans is op bepaalde misdaden. Op basis daarvan worden de routes van de politie aangestuurd. In een bepaalde wijk bleek het aantal fietsendiefstallen zeer sterk te dalen. Waren de politiepatrouilles daar zeer succesvol? Neen, uit onderzoek van Investico bleek dat in die wijk twee politiekantoren waren gesloten, mensen deden gewoon geen aangifte meer van fietsendiefstal. In de Verenigde Staten werd een algoritme gebruikt om het risico op recidief misdaadgedrag te schatten. Afro-Amerikanen die maar kleine misdaden hadden begaan, kregen een veel groter risico toegedicht dan blanken die zware feiten op hun kerfstok hadden. Toen journalisten achteraf onderzochten wie nu werkelijk opnieuw misdaden beging, bleek dat vaak niet te kloppen met de voorspellingen. De basis van het algoritme was in zekere mate racistisch. Het Michigan Unemployment Agency ging met een nieuw computersysteem van start om fraude met uitkeringen tegen te gaan. Het was zo streng dat zelfs het per vergissing fout schrijven van een naam of het ingeven van een verkeerde geboortedatum volstond om van fraude beschuldigd te worden. 20.000 werklozen, kwetsbare mensen die het geld hard nodig hadden, werden ten onrechte beschuldigd. Het is dus zeer belangrijk te weten welke data achter een algoritme zitten en op basis van welke vragen het is opgemaakt. Om bij het laatste voorbeeld te blijven: je zou ook een algoritme kunnen ontwikkelen dat mensen opspoort die recht hebben op een uitkering, maar ze nog niet hebben aangevraagd.’

Welk algoritme je bouwt en welke data je gebruikt, houdt dus een politieke en ethische keuze in.

‘Ja, het bouwen van algoritmes om uitkeringsfraude op te sporen bijvoorbeeld raakt aan fundamentele thema’s, aan de grondbeginselen van ons juridisch systeem en onze maatschappij. Je maakt bij voorbaat alle uitkeringstrekkers verdacht. Dat past naadloos in een lange geschiedenis van overheden die arme mensen in de gaten hielden, hen probeerden te controleren. Als we dat nu vastleggen in algoritmes, komen we er dan later nog wel weer uit? We moeten nu het bewustzijn ontwikkelen dat open standaarden en transparante algoritmes noodzakelijk zijn voor een democratie. Waar staat Europa in dit verhaal? Aan de ene kant zijn er de VS met een liberaal begrip van markt, waar bedrijven nu al deze functies overnemen en de algoritmes misschien bedrijfsgeheimen zijn. Aan de andere kant is er China waar de overheid bepaalt wat de bedrijven uit te voeren hebben. Europa moet zelf technologie ontwikkelen waar de *checks and balances* van democratie gewaarborgd blijven. Als we die technologieën gewoon kopen in de VS of China, kopen we ook het waardesysteem dat erin verankerd is.’

Hebt u de indruk dat beleidsmakers zich daar voldoende van bewust zijn?

‘Van twee soorten mensen ben ik een beetje bang: de ingenieur zonder begrip van de sociale wereld en de technocraat zonder begrip van de technologie. Maar als docent ben ik een onverbeterlijke optimist en ik denk dat onderwijs helpt. Er is nu veel meer aandacht voor de politieke en ethische aspecten van data. De bewustwording groeit. Steeds meer gemeenten, ministeries, organisaties, bedrijven vragen de hulp van onze Data School bij het kijken naar de ethische valkuilen van dataprojecten die ze willen opstarten. Ook burgemeesters, wethouders, raden moeten op dat punt veel competentier worden. Het is noodzakelijk dat ze basiskennis hebben van digitale en ethische vaardigheden. Een dataproject opstarten is een politieke beslissing: wat is het probleem en hoe kan een dataproject ons helpen om het op te lossen? Doen we het of doen we het niet? Hoe doen we het? Politici moeten ook de resultaten interpreteren, ze moeten dus weten wat er achter de data zit. Zij krijgen een rapport, een infografiek of een dashboard en moeten op basis daarvan een verantwoorde beslissing nemen. Met dezelfde data kunnen ze op verschillende manieren omgaan: ze kunnen ze gebruiken om een groep te stigmatiseren of net te ondersteunen. Neem het voorbeeld van een algoritme dat vrij goed voorspelt welke leerlingen de school vroegtijdig zullen verlaten. Het volstaat niet dat het aangeeft welke leerlingen je in de gaten moet houden, het gaat er vooral over wat je met de informatie doet. Welke kinderen zijn dat? Komen ze uit eenoudergezinnen? Hebben ze een migratieachtergrond? Is er een taalbarrière? Uit welke wijk komen ze? Wie spreekt de kinderen en hun ouders aan: iemand die aansluiting heeft met de leefwereld van de jongeren of een ambtenaar? Er ko-

men dus heel veel niet-data-aspecten bij kijken, die telkens een beslissing vragen. Belangrijk is ook dat een gemeente daar open over communiceert met de ouders, met de jongeren. Kan ik beslissen dat de data van mijn zoon niet voor die analyse gebruikt worden? Is er een opt-outfunctie? Hoe zit het met de beveiliging van de data? Stel dat zo’n dataset in de handen komt van een *recruiter*. Die heeft vast een model liggen waarmee hij kan bewijzen dat iemand die ooit in de risicogroep vroegtijdige schoolverlaters zat, in de toekomst slecht zal presteren op het werk. Hij wil die kandidaten misschien meteen uitsélectioneren. Al die aspecten moeten politici ook bekijken.’

Hoe helpt de Utrecht Data School gemeenten?

‘We hebben de Utrecht Data Ethics Decision Aid ontwikkeld, een impact assessment voor ethische valkuilen in dataprojecten. DEDA is een proces waarbij beleidsmakers, dataspecialisten, een functionaris gegevensbescherming, projectmanagers, inhoudelijk experts samen aan tafel zitten. Een van de eerste vragen is: gebruikt u een algoritme? De dataspecialist zegt: natuurlijk. De beleidsmaker valt vaak uit de lucht. De tweede vraag is of ze het algoritme kunnen uitleggen. De datamensen weten hoe dat ineenzit. De beleidsmaker denkt dat hij dat niet hoeft te kunnen, wat wil zeggen dat hij het ook niet kan uitleggen aan de raad, de bevolking, de media. Zo groeit het besef dat iedereen in de organisatie eigenlijk basiskennis moet hebben. Ook de wethouder moet weten hoe het algoritme werkt. Hij hoeft het niet te kunnen programmeren, maar hij moet wel de logica begrijpen, hij moet achter het beslissingsmodel van het algoritme kunnen staan en dat kunnen uitleggen. In de loop van de workshop komen allerlei vragen en valkuilen aan bod. Dat helpt om goede beslissingen te nemen, bijvoorbeeld dat ze niet alles hoeven te weten, dat ze zuinig moeten zijn op data en alleen die gegevens verzamelen die ze echt nodig hebben, dat ze die goed moeten beschermen enzovoort. Zo komen ze tot betere projecten waarvan het beslissingsproces gedocumenteerd wordt. Een kritisch publiek kan daar vragen over stellen. In Nederland zijn er al gemeenten die voor al hun dataprojecten een ethisch *impact assessment* of een ethische quick scan doen. Ook de Vereniging van Nederlandse Gemeenten gebruikt onze tool, iedereen kan hem downloaden op onze website.’

Hoe beginnen gemeenten met een databeleid?

‘Pilots zijn zeer belangrijk om met data te leren omgaan. In kleine projecten ontwikkel je vaardigheden en competenties. Zo werkte een team van de



Utrecht Data School samen met een Nederlandse gemeente om meer te weten te komen over wie de inwoners in de bijstand zijn. Tijdens het onderzoek doken allerlei moeilijkheden op: niet alle data bleken beschikbaar te zijn, sommige mochten niet gebruikt worden. Het college en de gemeenteraad beslisten het project voort te zetten om ervan te leren. Ze schakelden de privacy officer in, ze zorgden ervoor dat de verwerking van de data op een goede manier zou gebeuren en ze installeerden een manier van werken die ook voor toekomstige dataprojecten kan dienen. We zien ook in andere gemeenten een databrigade of een data-academie ontstaan. Dat zijn informele netwerken van geïnteresseerde medewerkers van alle niveaus en diensten die samen bespreken hoe ze beter met data kunnen werken. Dat is veelbelovend. Maar het kan niet zonder de leiding en de steun van het topmanagement en de beleidsmakers.'

Weten gemeenten over welke data ze beschikken?

'Nee, er zijn veel meer data beschikbaar dan de meeste gemeenten beseffen. In een rapport worden vaak bepaalde dingen uitgelicht, maar de dataset kan ook antwoorden geven op andere vragen. De eerste opdracht is dus andere vragen stellen aan de beschikbare datasets. Ten tweede moeten datasets kwalitatief beter worden. En dan is er ook nog het probleem van de standaarden voor data-uitwisseling. Elke gemeente verzamelt data op een verschillende manier. Het zou handig zijn Europese open standaarden te hebben, zodat de uitwisseling van data en kennis tussen gemeenten makkelijker wordt.'

Moeten inwoners bij dataprojecten betrokken worden?

'Dat hangt af van het soort project. Er zijn pogingen geweest om burgers te betrekken bij projecten

'Ook de wethouder moet weten hoe het algoritme werkt. Hij hoeft het niet te kunnen schrijven, hij moet wel de logica begrijpen, hij moet achter het beslissingsmodel van het algoritme kunnen staan en dat kunnen uitleggen.'

van datageletterdheid, maar die waren niet altijd succesvol en wie kwam opdagen was niet representatief voor de bevolking. Ik heb daar grote twijfels over, ik hou meer van een expertbenadering. Het gaat mij meer over open communicatie en accountability, verantwoording afleggen aan de raad van bestuur en aan de vierde macht. Net zoals een gemeente open is over haar verkeersplan, moet ze open zijn over de data die ze heeft, wat ze met data doet en wat ze ermee wil bereiken, zodat de raad en de media kritische vragen kunnen stellen. Als een gemeente voor een vroegtijdig schoolverlateralgoritme gaat, dan moet ze communiceren waarop het is gebaseerd, hoe het tot stand kwam, wat er met de data gebeurt enzovoort. Dan hoeft cocreatie niet. Maar in een ander project kan het wel nodig zijn. In Nederland is er een voorstel om slimme lantarenpalen met camera en microfoon in te zetten om vreemd gedrag op straat te kunnen monitoren. Een algoritme bepaalt dan of er sprake is van vreemd gedrag: rennen, roepen, vechten. Daarover moet de wijkbewoner wel kunnen meespreken, omdat het de persoonlijke levenssfeer raakt.'

Wat is de grootste opdracht voor lokale besturen?

'Ze mogen niet in de technocratische valkuil trappen. Ze mogen niet vergeten dat de data, de processen, de algoritmes moeten passen binnen ons democratisch stelsel, dat datapraktijken verantwoordelijkheid vragen.' ■

BART VAN MOERKERKE IS REDACTEUR VAN LOKAAL

www.dataschool.nl