

**BIG DATA**

Zelflerende algoritmen hebben niet de wijsheid in pacht, betoogt **Marc Steen**. Kennis van onderliggende processen is cruciaal om te begrijpen waar het misgaat.



Rapper Typhoon werd in 2016 in Zwolle staande gehouden op grond van etnische profilering. Zijn huidskleur, leeftijd en luxe auto vond de politie een verdachte combinatie. FOTO ANP

## Ook **algoritmen** maken fouten

**D**ankzij het debacle met Cambridge Analytica en Facebook zijn we ons meer bewust geworden van het belang van online privacy en dataprotectie. Om die te waarborgen, is in mei de AVG (Algemene Verordening Gegevensbescherming) van kracht geworden in Europa. Maar er is een veel groter probleem: het probleem van zelflerende algoritmen die op basis van stapels data allerlei conclusies trekken. Conclusies die we klakkeloos geloven, zonder uitleg. Hoogleraar psychologie Mark van Vugt roept in zijn wetenschapsrubriek op tot een debat daarover (*Trouw*, 2 juni). Dat debat is nodig en nuttig. Soms moet je een spade dieper graven.

Van Vugt noemt een algoritme dat op basis van foto's van gezichten criminelen van niet-criminelen zou onderscheiden. Twee hoogleraren van de Universiteit van Washington keken hier kritisch naar en leggen uit dat er een

*bias* (vertekening) in het algoritme is geslopen door de foto's waarmee het algoritme is getraind. Het algoritme is getraind met twee sets foto's: politiefoto's van veroordeelde criminelen, die nors kijken, en willekeurige foto's van professionele websites, waarop de mannen glimlachen.

Het algoritme heeft dus geleerd om onderscheid te maken tussen nors kijken en glimlachen. Niet om criminelen te herkennen. Zo zijn er talloze voorbeelden van algoritmen die conclusies trekken. In de rechtspraak, in het vinden van werk, in het verkrijgen van een lening of hypotheek, in het kiezen van medische behandelingen. En dat zonder uitleg over het onderliggende proces. Zo kunnen we dat proces niet goed bekritisieren. Dat moet beter.

### Mens en machine

Om beter te worden in het stellen van kritische vragen kunnen we drie begrippen gebruiken: *agency*, transparantie en

*accountability* van algoritmen.

*Agency* (het vermogen om autonoom te handelen) gaat over de verdeling van macht tussen mens en machine. Dit is een glijdende schaal. Aan de ene kant staat de mens die de machine gebruikt als gereedschap. Het is dan cruciaal dat hij begrijpt hoe de machine werkt. Aan de andere kant van de schaal staat de machine centraal en is de mens verworden tot een tandwiel. Hij leest op zijn scherm wat hij moet doen en begrijpt er weinig van. *Computer says no*. Het lijkt mij duidelijk dat we geen tandwielen willen worden.

### Begrijpen en uitleggen

Transparantie gaat over de mate waarin een algoritme inzichtelijk is. Kun je het begrijpen en uitleggen? Maar bedrijven of overheden willen of kunnen vaak niet vertellen hoe hun algoritmen werken. Dat kan zijn vanwege intellectueel eigendom of geheimhouding, maar het kan ook zijn omdat ze zelf evenmin be-

grijpen hoe hun (zelflerende) algoritme precies werkt.

*Accountability* gaat over het kunnen afleggen van verantwoording. Bijvoorbeeld over wat rechtvaardig of eerlijk is. Algoritmen gebruiken data uit de maatschappij. Daar komt nu eenmaal discriminatie voor en zo komt er (onbedoeld) discriminatie in de algoritmen terecht. Er zijn voorbeelden van discriminatie in de rechtspraak in de VS. Of over de mate van accuratesse. Algoritmen werken wel met grote getallen, maar maken fouten op individueel niveau. Zoals toen een politieagent rapper Typhoon onterecht staande hield (hij bleek niet-crimineel). In zulke gevallen heb je ook transparantie nodig.

Ik hoop dat we, als burgers en als consumenten, meer kritische vragen kunnen gaan stellen over algoritmen, en daarmee overheden en bedrijven dwingen om algoritmen te gebruiken die passen in een rechtvaardige samenleving.



### Marc Steen

senior onderzoeker bij TNO, o.a. in het onderzoeksprogramma Verantwoorde Waardecreatie met Big Data (VWData)

**DE VRAAG** van Monic Slingerland

Kijken boeren met oogkleppen op naar hun toekomst?