

## ARTIKELEN

### THEMA-ARTIKEL

# Transparantie van algoritmen: naar een empirische onderzoeksagenda<sup>\*</sup>

*Haiko van der Voort & Joanna Strycharz*

Er is veel aandacht voor de manier waarop overheden hun analyses uitvoeren. Big data, artificiële intelligentie en *internet of things* geven toegang tot data en veel mogelijkheden om de analyses te optimaliseren. Publicisten en beleidsmakers waren de nodige jaren enthousiast over de mogelijkheden van de nieuwe technologieën en analysemethoden (Höchtel, Parycek, & Schollhammer, 2016; Janssen & Kuk, 2016). Nu wordt er veel gepubliceerd over de gevaren (Coglianese & Lehr, 2019; Redden, 2018).

Algoritmen, ofwel de formules op basis waarvan we computers analyses laten uitvoeren, beschikken over een air van neutraliteit. Maar analyses op basis van algoritmen blijken niet altijd neutraal, maar afhankelijk van de kwaliteit van de data en natuurlijk de kwaliteit van het algoritme zelf. Beide vergen een menselijke inbreng en zijn daarmee vatbaar voor fouten of bias. Bovendien vinden er vaak interpretatieslagen plaats naar aanleiding van het werk van de computer, alvorens tot een voor besluitvormers werkbare analyse kan worden gekomen. Ook deze zijn kwetsbaar en afhankelijk van menselijke input.

Zo lijkt artificiële intelligentie net zo feilbaar als de menselijke intelligentie die we al hadden, ook al omdat het niet zonder die menselijke intelligentie kan. AI blijkt 'slechts' een veelbelovende mechaniek in het analyseproces voor beleid. AI heeft wel potentieel forse effecten op personen. Wat als AI wordt gebruikt om kandidaten te selecteren voor een functie? Of voor het detecteren van fraude? Of voor het inzetten van de politie? Dan wordt de levenssfeer van individuen vrij direct beïnvloed op basis van feilbare methoden. AI kan daarmee op gespannen voet staan met individuele grondrechten (Gerards, Nehmelman, & Vetzo, 2018). De onrust wordt verder versterkt door het *unheimische* gevoel dat mensen geen vat hebben op algoritmen. Concreet begrijpen weinigen hoe computers tot hun output komen. Wanneer worden er algoritmen gebruikt? Wat zijn ze? Waar houdt de computer op en begint de mens?

<sup>\*</sup> Dr. H.G. van der Voort is universitair docent Organisatie & Governance aan de TU Delft, Faculteit Techniek, Bestuur en Management. J. Strycharz, Msc is universitair docent Persuasive Communication aan de Universiteit van Amsterdam, Faculteit Maatschappij- en Gedragwetenschappen.

Haiko van der Voort & Joanna Strycharz

Transparantie is de meest gehoorde belofte in deze context, zoals ook uiteengezet in de Kamerbrief van Verhoeven (D66) in 2018.<sup>1</sup> Maar het lijkt lastig om verder te komen dan het oproepen tot transparantie. In dit themanummer exploreren we de praktische en beleidsmatige complexiteiten van transparantie van algoritmen. We zien daarbij grofweg twee manieren om naar transparantie van algoritmen te kijken:

*Transparantie als oplossing voor een probleem.* Er is een probleem geconstateerd, namelijk een feilbaar proces dat impact heeft op individuen. Transparantie is vervolgens een logische oplossing in het publieke domein. Transparantie is immers een voorwaarde voor verantwoording, hetgeen essentieel is voor publieke besluitvorming. De oplossing moet vervolgens uitgewerkt worden. Transparantie organiseren kent wel wat haken en ogen. Eenvoudigweg een algoritme publiceren lijkt schijntransparantie. Immers, wie begrijpt het? Een volgende vraag is wat precies transparant gemaakt moet worden en waarom (Leenes, 2016). Dergelijke vragen zijn relevant wanneer we niet precies zien hoe menselijke en automatische intelligentie op elkaar ingrijpen. Bovendien: wellicht waren besluiten ook zonder algoritmen niet altijd transparant.

Stephan Grimmelikhuijsen en Albert Meijer richten zich in dit themanummer op het vertrouwen van de burger in en naar aanleiding van algoritmische besluitvorming. Teneinde dit vertrouwen te behouden is verantwoord algoritmegebruik volgens hen noodzakelijk. De auteurs specificeren dit idee in waardengevoeligheid en transparantie. Aan de hand van klassieke literatuur over informatisering in het openbaar bestuur vullen zij deze principes in.

Hans de Bruijn, Marijn Janssen en Martijn Warnier geven een kritische analyse van een veelbelovend concept van transparantie: Explainable Artificial Intelligence (XAI). Volgens dit idee zou verantwoording van overheidsbesluiten bevorderd worden indien algoritmen ‘uit te leggen’ zouden zijn. De gedachte gaat aan de complexiteit van overheidsbesluitvorming voorbij, zo lijken de auteurs te stellen aan de hand van zeven beperkingen van XAI. Ze sluiten af van met evenzoveel strategieën om ‘het kind niet met het badwater weg te gooien’.

*Transparantie als waarde.* Transparantie is essentieel in een democratie, omdat het noodzakelijk is voor publieke verantwoording. Als zodanig is het een waarde. Deze gedachte nodigt uit naar een bredere kijk: publieke besluitvorming is vaak een afweging tussen meerdere waarden. Minister Dekker gaf in reactie op de motie-Verhoeven al aan dat transparantie ten koste kan gaan van privacy en innovatie.<sup>2</sup> Een andere afweging betreft transparantie en de zorgvuldigheid van de analyse. Immers, transparantie is gediend bij eenvoudige algoritmen, maar juist eenvoudige algoritmen respecteren de complexiteit van deze wereld minder. En bovendien: wie maakt deze ingewikkelde afwegingen, als we weten dat het analyseproces van data genereren via interpretatie naar besluitvorming lang is en

1 Kamerstukken II 2017/18, 32761, nr. 117.

2 Handelingen II 2017/18, nr. 90, item 3, p. 4-5.

bevolkt wordt door meerdere experts en beleidsmakers. Welke prikkels ondervindt deze?

Oskar Gstrein en Andrej Zwitter onderscheiden ethische, juridische en sociale uitdagingen van *Predictive Policing* (PP) systemen met behulp van voorbeelden en klassieke literatuur. Hoe verhouden praktijken van PP zich tot de democratische rechtsstaat, vanuit verschillende interpretaties?

Joanna Strycharz, Bauke Steenhuisen en Haiko van der Voort problematiseren transparantie van algoritmen als waarde. Al snel wordt duidelijk dat er meerdere waarden in het spel zijn. Waardenafwegingen zijn noodzakelijk, maar hoe worden die afwegingen gemaakt in een organisatie? Op basis van theorie over waardenconflict wijzen zij daarnaast op de moeilijkheid om transparantie af te wegen tegen andere waarden.

Bij elkaar leveren de bijdragen een beeld van de complexiteit van transparantie van algoritmen vanuit meerdere invalshoeken. Gstrein en Zwitter zetten dilemma's en uitdagingen van algoritmisering uiteen vanuit een ethische en een juridische invalshoek. Grimmelikhuijsen en Meijer gebruiken dezelfde invalshoek, maar werken toe naar principes van verantwoorde algoritmisering met behulp van literatuur over informatisering in het openbaar bestuur. De Bruijn en zijn co-auteurs schrijven vanuit een technische en een bestuurskundige invalshoek. Zij zetten de uitlegbaarheid van algoritmen in een organisatorische context. Strycharz en haar co-auteurs zetten ten slotte waardenafwegingen in een organisatorische context.

De nadruk op uitdagingen en complexiteiten leiden in opzet niet tot een deprimerend themanummer. De auteurs geven vele aanbevelingen. Bovendien leiden de complexiteiten tot empirische onderzoeksvragen, die beantwoord zullen moeten worden om betekenisvol over transparantie van algoritmen te kunnen praten. Gstrein en Zwitter stellen voor om tijdens het ontwerpproces van algoritmen een open en brede discussie over het doel en de doelstellingen te voeren. Grimmelikhuijsen en Meijer stellen onderzoek voor naar de vaststelling van de verantwoordelijkheid voor algoritmiseringsprocessen, die immers al gauw bij verschillende partijen ligt. Zij suggereren ook etnografisch onderzoek bij ambtenaren en onderzoek naar hoe burgers overheidsbesluiten ervaren. Strycharz, Steenhuisen en Van der Voort stellen onderzoek voor naar de professionalisering van data-analyse, inclusief het claimen van autonomie door data scientists. Zij schetsen ten slotte een methodologische uitdaging hoe waardenafwegingen vast te stellen indien waardenafwegingen impliciet en ongelijktijdig plaatsvinden.

## Literatuur

- Coglianesi, C., & Lehr, D. (2019). Transparency and algorithmic governance. *Administrative Law Review*, 71.
- Gerards, J., Nehmelman, R., & Vetzo, M. (2018). *Algoritmes en grondrechten*. Den Haag: Boom Juridische uitgever.

Haiko van der Voort & Joanna Strycharz

- Höchtel, J., Parycek, & Schollhammer. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing*, 26(1-2), 147.
- Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 3-13. <https://doi.org/10.1080/10919392.2015.1124005>
- Leenes, R. (2016). De voorspellende overheid. *Bestuurskunde*, 25(1), 38-43.
- Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*. <https://doi.org/https://doi.org/10.1177/2053951718809145>