

Eerlijkheid door archivering

De bijdrage van archivering aan eerlijk algoritmegebruik door gemeentelijke overheden

Naam: Tobias van der Knaap

Studentnummer: 10562168

Begeleider UvA: Prof. dr. K.J.P.F.M. (Charles) Jeurgens

Tweede lezer: mr. dr. T.A. (Tjeerd) Schiphof

Begeleiders CBS: dr. B. (Barteld) Braaksma

A. (Anna) Mitriaieva MSc

Masterscriptie Erfgoedstudies: Archival and Information Studies (dual)

Datum: 31 augustus 2020

Aantal woorden: 20.708



**Centraal Bureau
voor de Statistiek**



UNIVERSITEIT VAN AMSTERDAM

Abstract

With rapid developments in Artificial Intelligence (AI), local governments in the Netherlands are increasingly becoming interested in harnessing this new technology to improve their own work processes. This thesis describes the results of a research project carried out in collaboration with Statistics Netherlands (CBS), with the goal of discovering how archival and information policies can contribute to the process of securing fairness, transparency, explainability and responsibility in algorithm usage by Dutch municipal governments. Based on academic literature from the fields of philosophy of technology and archival science, a study of existing policy and regulation frameworks and interviews with civil servants, this thesis describes the current situation regarding municipal archiving and information management, as well as the state of algorithm usage in municipal governments. It then suggests ways to improve this situation, through the overarching principle of ensuring the traceability of the process of governmental decision-making via thorough and complete archiving of the decision-making process in all its stages.

Keywords: Artificial Intelligence, Algorithms, Governance, Archival Policy, Information Management

Inhoud

Abstract	2
Inleiding	4
Hoofdstuk 1: Literatuur	10
Hoofdstuk 2: Regulerende en beleidskaders	20
Hoofdstuk 3: Interviews	31
Hoofdstuk 4: Synthese en aanbevelingen	44
De aanbevelingen opgesomd	54
Conclusie	55
Bijlage 1: Literatuurlijst	60
Bijlage 2: <i>interview guides</i>	65
Interview guide medewerkers algoritmeprojecten	65
Interview guide archief- en informatieprofessionals	66

Inleiding

In een video getiteld 'Data' brengt de Britse acteur en filosoof Oliver Thorn lastige kwesties rond technologiegebruik aan het licht.¹ De video toont twee personages, beiden gespeeld door Thorn: een jonge clubbezoeker en een uitsmijter, die het legitimatiebewijs van de bezoeker wil scannen voor hij deze naar binnen laat. De bezoeker stelt hier kritische vragen bij. Waarom is het nodig dat zijn legitimatie gescand wordt? Wat gebeurt er met de daardoor verkregen informatie? Waarom gaat de uitsmijter hier zo makkelijk mee akkoord? Door slimme montage toont de video verschillende versies van deze socratische dialoog: er wordt regelmatig teruggespoeld naar een eerder punt in het gesprek, waarin de uitsmijter een ander antwoord op de vraag probeert en zo het gesprek hoopt te kunnen 'winnen'.

De dialoog volgt zo verschillende paden, waarin talloze aspecten van datagedreven surveillance aan bod komen. Een belangrijke rode draad hierin is de rol van de uitsmijter als degene die de technologie hanteert. Telkens weer lukt het de bezoeker te laten zien dat er aspecten zijn waar de uitsmijter onvoldoende over heeft nagedacht en dat het scannen van een legitimatiebewijs meer is dan alleen een simpele technische handeling. Thorn kiest er niet voor niets voor om de uitsmijter geen replica van een echte scanner in zijn hand te geven, maar een hamer. Misschien kwam deze keuze deels voort uit het niet beschikbaar zijn van een goed rekwisiet, maar het is ook een metafoor voor hoe de scanner gebruikt en gezien wordt: als een simpel stuk gereedschap, gemaakt om mensen te helpen bij het uitvoeren van een taak.

Het voornaamste punt van de video, is te laten zien dat er meer achter moderne technologieën zit dan je intuïtief zou verwachten bij het hanteren van die technologie. Deze scriptie biedt een vergelijkbaar perspectief op een specifiekere casus: het gebruik van algoritmen door Nederlandse gemeenten. Een algoritme is, op de meest basale manier omschreven, niet meer dan een geprogrammeerde reeks instructies om ingevoerde data te bewerken tot een eindproduct.² Op het eerste gezicht klinkt dit niet als hele spectaculaire technologie: de populaire vergelijking is dat een algoritme niet wezenlijk anders is dan een cakerecept, wat immers ook bestaat uit input (ingrediënten), een reeks instructies en output (de cake). De afgelopen jaren zijn er echter grote stappen gemaakt in het ontwikkelen van algoritmen, waardoor de instructies steeds complexer en uitgebreider zijn geworden en er steeds meer toepassingen mogelijk zijn. Er is vooral belangstelling

¹ Philosophy Tube, *Data*. Gemaakt door Oliver Thorn. 31 januari 2020, te raadplegen op: <https://www.youtube.com/watch?v=fCUTX1jurJ4>. Laatst geraadpleegd op 31 juli 2020.

² M. Craglia (Ed.) e.a., *Artificial Intelligence – A European Perspective* (Luxemburg 2018) 57-59. Te raadplegen op: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf>. Laatst geraadpleegd op 19 maart 2020.

voor de mogelijkheden van algoritmen om patronen in grote hoeveelheden gegevens te vinden en zo tot nieuwe inzichten te komen.³

Veel van de grote ontwikkelingen op het gebied van algoritmen komen uit de wetenschap en uit de private sector, maar overheden tonen er de laatste jaren ook steeds meer belangstelling voor. Onderzoeken van TNO en het Centraal Bureau voor de Statistiek (CBS) lieten recent zien dat diverse overheden, van gemeenten tot landelijk, gebruik maken van algoritmen, vooral in het sociaal domein en op het gebied van veiligheid.⁴ Dit past in een bredere context van toenemende digitalisering van overheden, die steeds meer volledig digitaal werken en gebruik proberen te maken van nieuwe technologieën.⁵ Met dit gebruik gaan echter ook risico's gepaard, zoals eerder dit jaar bleek uit de geruchtmakende uitspraak in de rechtszaak rondom het Systeem Risico Indicatie (SyRI), gebruikt bij de opsporing van bijstandsfraude.⁶ SyRI was volgens de rechter onvoldoende controleerbaar en maakte een onevenredig grote inbreuk op de privacy van burgers, waardoor het niet meer gebruikt mag worden. De grootschalige data-analyses die algoritmen uitvoeren, maken veel nieuwe dingen mogelijk, maar het is dus de vraag in hoeverre al die toepassingen wenselijk zijn en, zo nee, hoe ze wel wenselijk ingericht kunnen worden.

Het CBS voerde in dit kader het project Eerlijke Algoritmen uit, als onderdeel van een groter samenwerkingsverband getiteld *AI met Impact*.⁷ Hierin werd gewerkt aan manieren om AI-gebruik bij Nederlandse gemeenten eerlijk en transparant te maken. Deze scriptie is het resultaat van onderzoek uitgevoerd in het kader van dit project bij het CBS, met de specifieke invalshoek om de rol van archivering hierbij te onderzoeken. De reden hiervoor is de traditionele rol van overheidsarchieven als instrument van verantwoording voor overheden.⁸ De hoofdvraag die centraal staat in dit onderzoek is dan ook: hoe kan de borging van transparantie, uitlegbaarheid, verantwoording en eerlijkheid bij het gebruik van algoritmes door Nederlandse gemeenten verbeterd worden door

³ Ethem Alpaydin, *Introduction to Machine Learning* (Cambridge MA 2010), 1.

⁴ Anne Fleur van Veenstra e.a., *Quick scan AI in de publieke dienstverlening* (TNO, 2019). Te raadplegen op: <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/04/08/quick-scan-in-de-publieke-dienstverlening/TNO+-+AI+in+de+overheid+eindpresentatie+08042019.pdf>. Laatst geraadpleegd op 29 juli 2020; S. Doove & D. Otten, *Verkenkend onderzoek naar het gebruik van algoritmen binnen overheidsorganisaties*. (Den Haag: CBS 2018). Te raadplegen op: <https://www.cbs.nl/nl-nl/maatwerk/2018/48/gebruik-van-algoritmen-door-overheidsorganisaties>.

⁵ Zie bijvoorbeeld deze website van de Rijksoverheid, waarop de wens wordt uitgesproken om voorop te lopen 'bij het gebruik van nieuwe digitale technieken.': <https://www.digitaleoverheid.nl/dossiers/nieuwe-technologieen-data-en-ethiek/>. Laatst geraadpleegd op 16 augustus 2020.

⁶ Zie bijvoorbeeld: Camil Driessen, 'Rechtbank: overheid moet stoppen met fraudesysteem', website *NRC Handelsblad*, 5 februari 2020. Te raadplegen op: <https://www.nrc.nl/nieuws/2020/02/05/rechtbank-overheid-moet-stoppen-met-fraudesysteem-a3989325>. Laatst geraadpleegd op 31 juli 2020.

⁷ Zie de website van partnerorganisatie CodeForNL voor meer informatie: <https://www.codefor.nl/ai-met-impact/>.

⁸ Ministerie van Onderwijs, Cultuur en Wetenschap, *Van Archiefwet 1995 naar Archiefwet 202x: de belangrijkste wijzigingen van de nieuwe Archiefwet 202x, met korte toelichting*. (Den Haag 2019), te raadplegen op: <https://www.internetconsultatie.nl/archiefwet/document/5165>, 1-2.

beleidsveranderingen op het gebied van archivering en informatiebeheer? Om volledig aan te sluiten bij het project Eerlijke Algoritmen, richt deze scriptie zich uitsluitend op gemeenten, ook al worden algoritmen inmiddels in vrijwel alle lagen van de overheid toegepast.⁹

Enige toelichting op deze vier criteria is wel even van belang. Eerlijkheid stond centraal in het CBS-project, aangezien het een thema is waar veel maatschappelijk belang aan wordt gehecht. Binnen het project Eerlijke Algoritmen wordt eerlijkheid gedefinieerd als de afwezigheid van ongewenste discriminatie.¹⁰ De zorgen over oneerlijke behandeling van individuen door algoritmen zijn wijdverbreid. Cathy O’Neill schrijft in haar bestseller *Weapons of Math Destruction* over de risico’s die massale analyse en classificatie door algoritmen met zich meebrengen als er onvoldoende oog is voor de maatschappelijke omstandigheden waarin wordt gewerkt.¹¹ De centrale stelling van het werk van O’Neill is dat deze technologieën op deze manier alleen maar bestaande ongelijkheden in stand houden of zelfs verergeren.¹² Daarom is het ook belangrijk om te kijken hoe ongewenste discriminatie kan worden tegengegaan – oftewel hoe eerlijkheid gebord kan worden.

De plek van verantwoording is vooral ontleend aan de eerder genoemde rol van archivering in het veiligstellen van verantwoording. Door het bewaren van gearchiveerde stukken, kan het handelen van overheden immers onderzocht en gecontroleerd worden, waardoor de overheid verantwoordelijk kan worden gehouden. Transparantie raakt hieraan, aangezien dit gaat om openheid van zaken geven, veelal door documenten openbaar te maken. Over dit laatste was ook in het SyRI-proces het nodige te doen, aangezien de staat geen inhoudelijke details over de werking van het programma wilde geven. Uitlegbaarheid is een wat nieuwere term, maar een die steeds meer gebruikt wordt wanneer het om algoritmegebruik gaat, bijvoorbeeld door de Europese Commissie en het ministerie van Justitie en Veiligheid (zie hoofdstuk 2). Algoritmen zijn inmiddels zodanig complex geworden, dat het risico bestaat dat ze in zogeheten *black boxes* veranderen: ondoorzichtige ‘zwarte dozen’ die output leveren waarvan volstrekt onduidelijk is hoe die tot stand is gekomen. Met uitlegbaarheid kan dit mogelijk voorkomen worden.

Deze vier kwalitatieve criteria staan centraal in dit onderzoek. Gezamenlijk vormen ze het kader voor wenselijke algoritmegebruik binnen een democratische rechtsstaat: als het algoritmegebruik voldoende uitlegbaar en transparant is, is het inzichtelijk voor burgers. Dit draagt bij aan het borgen van verantwoording, omdat burgers beter geïnformeerd zijn en de overheid

⁹ Van Veenstra, *Quick Scan*; Doove en Otten, *Verkenkend onderzoek*.

¹⁰ Rik Helwegen en Barteld Braaksma, *Fair Algorithms in Context* (Den Haag 2020), 13. Te raadplegen op: <https://www.cbs.nl/nl-nl/over-ons/innovatie/project/fair-algorithms-in-context>. Laatst geraadpleegd op 10 augustus 2020.

¹¹ Cathy O’Neill, *Weapons of Math Destruction. How big data increases inequality and threatens democracy* (Londen 2017) 155-160.

¹² Ibidem, 11-13.

daardoor gemakkelijker ter verantwoording kunnen roepen voor de handelingen die door of met ondersteuning van algoritmen zijn uitgevoerd. Dit zorgt er verder voor dat er betere controle is of er sprake is van eerlijk handelen door de overheid. Zo ondersteunen en versterken deze criteria elkaar. In dit onderzoek vormen ze het kader voor het analyseren van academische literatuur (in hoofdstuk 1), regelgevende beleidskaders (in hoofdstuk 2) en interviews met mensen uit het veld (in hoofdstuk 3), om in kaart te brengen wat de bestaande situatie is rondom het borgen van deze kwaliteitscriteria. In de aanbevelingen van hoofdstuk 4 wordt dan vervolgens bekeken hoe deze kan worden verbeterd met behulp van archivering en informatiebeheer.

Literatuurbespreking en theoretische inkadering van het probleem

In de wetenschappelijke literatuur is ook steeds meer aandacht voor de maatschappelijke rol van algoritmen. Het bredere filosofische kader voor dit onderzoek bestaat vooral uit het werk van de Twentse hoogleraar Peter-Paul Verbeek, die zich als techniekfilosoof bezighoudt met vraagstukken over de verhouding tussen mens en techniek en de morele kwesties die hierbij meespelen. Verbeek ziet mens en techniek als onlosmakelijk verbonden en stelt dan ook dat het niet mogelijk voor mensen is om zich aan de invloed van technologie te onttrekken.¹³ Dit betekent echter niet dat we de invloed maar gewoon moeten accepteren en geen poging kunnen doen om deze bij te sturen. Voor Verbeek is dit namelijk de kern van de discussie die over technologie gevoerd zou moeten worden: hoe kunnen bestaande technologieën toegepast worden op een manier die wenselijk is?¹⁴

Buiten het werk van O'Neill zijn er tal van andere academici die zich specifiek met de maatschappelijke kanten van grootschalige algoritmische data-analyse bezighouden. Een voorbeeld hier van is een artikel van Lucas Introna, waarin het gaat om de problemen bij het in kaart brengen van de productie van software waaraan door meerdere mensen en teams gewerkt is.¹⁵ Introna concludeert dat alleen de code onvoldoende inzicht geeft en dat het proces en de context waarmee deze tot stand is gekomen minstens zo belangrijk is om te zien op welke manier en met welke intenties het programma gemaakt is. Ditzelfde geldt voor de datasets die gebruikt worden als input voor algoritmen, aldus danah boyd en Kate Crawford.¹⁶ Zij constateren dat er vaak onvoldoende oog

¹³ Peter-Paul Verbeek, 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research'. Gepubliceerd in: J.K. Berg O. Friis and Robert C. Crease, *Technoscience and Postphenomenology: The Manhattan Papers* (London 2016), pp. 189-204. Geraadpleegd op: <https://research.utwente.nl/en/publications/toward-a-theory-of-technological-mediation-a-program-for-postphen>.

¹⁴ Peter-Paul Verbeek, *Op de vleugels van Icarus. Hoe techniek en moraal met elkaar meebewegen* (Rotterdam 2014) 8.

¹⁵ Lucas D. Introna, 'Algorithms, Governance, and Governmentality: On Governing Academic Writing', in: *Science, Technology & Human Values* 41 (2016), 17-49.

¹⁶ danah boyd & Kate Crawford, 'Critical Questions for BigData', in: *Information, Communication & Society* 15:5 (2012), 662-679.

is voor de herkomst van datasets en voor de aanvankelijk onzichtbare selectieprocessen die voorafgaan aan de totstandkoming van een dataset.

Binnen de archiefwetenschap wordt ook steeds meer over algoritmen geschreven. In de Nederlandse context is dan vooral de bundel *Archives in Liquid Times* relevant, waarin diverse auteurs schrijven over archieven en het digitale tijdperk, waarin veel principes uit het papieren tijdperk op nieuw overdacht moeten worden.¹⁷ Eric Ketelaar schrijft in zijn voorwoord voor de bundel dat theoretische inkadering en bewustzijn van de grotere kwesties essentieel zijn voor het oplossen van de praktische vraagstukken waar veel mensen en organisaties in het digitale tijdperk mee geconfronteerd worden als het om archivering gaat.¹⁸ Een nog recentere bijdrage aan de discussie komt van Herbjørn Andresen, die een belangrijke rol ziet weggelegd voor archivering en informatiebeheer als het gaat om de verantwoording van algoritmegebruik.¹⁹

Overzicht

Er is dus duidelijk ruimte in de literatuur voor een studie als deze, zeker omdat er weinig praktijkonderzoek gedaan is naar de mogelijke rol van archivering in algoritmegebruik. Het blijft wel geregeld een breed onderzoek, aangezien er ook gekeken wordt naar archivering als geheel en naar de brede inzetbaarheid en impact die algoritmen kunnen hebben. Ter ondersteuning van het praktische onderzoek begint deze studie in hoofdstuk 1 met een uitgebreidere theoretische inkadering van de problematiek, waarin de vraag centraal staat hoe technologieën in verhouding staan tot hun politiek-maatschappelijke context. Dit geeft een brede filosofische en theoretische context voor de rest van het onderzoek. Dit contextualiseert de uiteindelijke analyse en beantwoording van de hoofdvraag, door een raamwerk te creëren waarbinnen de politiek-maatschappelijke kant van algoritmegebruik door de overheid gezien kan worden.

Hoofdstuk 2 gaat al wat meer richting de praktijk, door te kijken welke regelgevende en beleidskaders er zijn rondom overheidsinformatiebeheer en algoritmen, en hoe transparantie, uitlegbaarheid, verantwoording en eerlijkheid hierin terugkomen. Dit geeft een beter begrip van de situatie in de praktijk, een beeld van de tekortkomingen die er momenteel nog zijn in de wet- en regelgeving en laat zien hoe de genoemde criteria al verwerkt zijn in de wettelijke- en beleidskaders. Dit zorgt ervoor dat er een duidelijker beeld ontstaat van de huidige toestand wat betreft wet- en regelgeving rondom zowel overheidsinformatiebeheer als algoritmegebruik, waardoor beter te

¹⁷ Frans Smit, Arnoud Glaudemans en Rienk Jonker ed, *Archives in Liquid Times* (Den Haag 2017).

¹⁸ Eric Ketelaar, *Foreword*, in: Frans Smit, Arnoud Glaudemans en Rienk Jonker ed, *Archives in Liquid Times* (Den Haag 2017), VII.

¹⁹ Herbjørn Andresen, 'A discussion frame for explaining records that are based on algorithmic output', in: *Records Management Journal* 30:2 (2020), <https://doi.org/10.1108/RMJ-04-2019-0019>.

formuleren is hoe archivering een aanvulling kan zijn op de bestaande wet- en regelgeving en dus kan bijdragen aan eerlijk, transparant, uitlegbaar en verantwoord algoritmegebruik.

Voor hoofdstuk 3 zijn interviews afgenomen om meer informatie te verzamelen over de huidige situatie in de praktijk op het gebied van archivering en informatiebeheer binnen gemeenten en hoe er momenteel met algoritmen wordt gewerkt binnen gemeenten. Alle geïnterviewden waren werkzaam bij een gemeente, of bij een archiefdienst die aan een of meerdere gemeenten diensten verleend. De bedoeling van de interviews was om mogelijke problemen in het dagelijks werk in kaart te brengen, vooral waar het ging om archivering en informatiebeheer en digitaal werken. Hoewel veel gemeenten de overgang naar digitaal werken al enige tijd hebben gemaakt, zijn er op dit gebied immers nog tal van onopgeloste vraagstukken. Door het te combineren met gesprekken met medewerkers van projecten waarbij algoritmen betrokken zijn, kon gekeken worden naar hoe deze problemen zich specifiek manifesteren als het om algoritmen gaat. Het doel hiervan is wederom om bij te dragen aan een duidelijker beeld van de bestaande situatie en de tekortkomingen daarvan, zodat beter beschreven kan worden hoe archivering kan helen in het tegengaan van deze tekortkomingen.

Hoofdstuk 4, tot slot, bevat aanbevelingen geformuleerd op basis van het onderzoek dat voor de eerste drie hoofdstukken is uitgevoerd. De aanbevelingen richten zich vooral op informatiebeleid en –cultuur, op basis van actuele archieftheoretische inzichten, en meer concrete maatregelen om de juiste contexten en de uitlegbaarheid van het algoritme veilig te stellen. Hoewel de aanbevelingen in algemene zin waarschijnlijk breder toepasbaar zijn, zijn ze gebaseerd op en bedoeld voor de situatie in Nederlandse gemeenten. Verder geven ze een praktische invulling aan de beantwoording van de hoofdvraag, met adviezen hoe er in de praktijk gezorgd kan worden dat bij algoritmegebruik aan de vier genoemde criteria voldoet. De conclusie biedt tot slot een bredere, overkoepelende analyse van de maatschappelijke problemen die kunnen ontstaan door algoritmegebruik waarbij onvoldoende oog is voor de reeds genoemde vier criteria en wat er op maatschappelijk niveau nodig is om te zorgen voor een wenselijke implementatie van de technologie.

Hoofdstuk 1: Literatuur

Mag een verdachte zijn proces in vrijheid afwachten, of moet dat in de cel? Voor het beantwoorden van deze vraag hebben Amerikaanse rechters al enige tijd de beschikking over het programma COMPAS, dat op basis van historische gegevens aan verdachten een score op een schaal van 1 tot 10 toekent, die aangeeft hoe groot het risico is dat de betreffende verdachte tijdens een eventuele vrijlating weer een strafbaar feit zal begaan. Onlangs publiceerde de *MIT Technology Review* een interactief artikel over COMPAS, waarin lezers zelf aan de slag kunnen met het verschuiven van de drempelwaardes voor vrijlaten of opsluiten.²⁰ Het artikel laat zien hoe problematisch het werken met een dergelijk programma is, in een maatschappij waarin eeuwen aan systematisch racisme ervoor hebben gezorgd dat de historische gegevens waarop het algoritme getraind is ernstig bevooroordeeld zijn, wat leidt tot een ongelijke behandeling van met name Afro-Amerikaanse verdachten. Daarnaast zet het de lezer vooral aan het denken over wat rechtvaardige of eerlijke drempelwaardes zijn en hoe belangrijk dergelijke toepassingen kunnen zijn, niet alleen voor de verdachte, maar ook voor het maatschappelijk rechtvaardigheidsgevoel.

Het artikel over COMPAS laat zien dat het gebruik van algoritmen raakt aan fundamentele maatschappelijke kwesties als gelijke behandeling, rechtvaardigheid en institutioneel racisme. Tegelijkertijd wordt hierdoor ook zichtbaar hoe het algoritme is beïnvloed door de maatschappelijke context waarin het is ontwikkeld en wordt gebruikt: door historische data over te nemen uit een rechtssysteem dat bevooroordeeld is tegen sommige bevolkingsgroepen, komen deze zelfde vooroordelen in het algoritme terecht. In dit hoofdstuk wordt, aan de hand van wetenschappelijke literatuur, deze verstrengeling tussen techniek en maatschappij verder geanalyseerd. De vraag die hierbij centraal staat, is hoe technologie, en algoritmen in het bijzonder, in verhouding staan tot hun politiek-maatschappelijke context. Dit schetst een breder theoretisch en filosofisch kader ter ondersteuning van de analyse van de beleidsstukken en de interviews en draagt als zodanig bij aan de beantwoording van de hoofdvraag. Allereerst worden ideeën uit de techniekfilosofie besproken, met name het werk van Peter-Paul Verbeek, om een beeld te krijgen van het huidige denken over de relatie tussen mensheid en technologie. Vervolgens worden enkele studies besproken die de voor dit onderzoek belangrijkste nieuwe technologische ontwikkelingen analyseren: *machine learning* algoritmes en big data.

²⁰ Karen Hao en Jonathan Stray: 'Can you make AI fairer than a judge?', in: *MIT Technology Review*: <https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>, gepubliceerd 17 oktober 2019, laatst geraadpleegd op 4 januari 2020.

Techniekfilosofie

Een belangrijk denker op het gebied van de relatie tussen mens en technologie, is de Twentse hoogleraar Peter-Paul Verbeek. Centraal in het werk van Verbeek staat de rol van technologie als bemiddelaars ('mediators') van menselijke ervaringen en praktijken. In een in 2016 gepubliceerd artikel pleit hij ervoor om bestaande technologieën als uitgangspunt te nemen voor filosofische analyse, en daarbij ook vooral te kijken naar hoe mensen betekenis geven aan de bemiddeling van die technologieën, door een combinatie van filosofische en empirische methoden.²¹ In hetzelfde artikel bekritiseert Verbeek ook de Actor-Network Theory van Bruno Latour, omdat deze onvoldoende ruimte zou laten voor wat Verbeek de hermeneutische kant van bemiddelingen noemt: de manier waarop technologie actieve invloed uitoefent op hoe mensen de wereld interpreteren.

In een eerder artikel, geschreven met Asle Kiran, stelt Verbeek verder dat het weinig zinvol is om mens en technologie als gescheiden, tegengestelde categorieën te zien, omdat dit ervoor zorgt dat onderlinge verbindingen en relaties onzichtbaar blijven.²² De auteurs schrijven, op basis van Heideggers ideeën over de verhoudingen tussen mens en gereedschap, dat technologieën mensen en hun levenswereld constitueren door hoe zij door mensen gehanteerd worden: 'Using technologies means that we are in our surroundings in specific ways, relating ourselves to the aspects of the world that technologies help to accentuate and stand out. Technologies co-shape the appearance of the world; we do not just see a world, the world appears to us in certain ways: technologies structure and organize the world.'²³

Dit idee van een onlosmakelijke band tussen mens en technologie is een belangrijk onderdeel van het werk van Verbeek. Hij geeft hiermee ook een grote rol aan technologie, zoals te zien is in het bovenstaande citaat. Het idee van technologie als een extensie van het menselijk lichaam wijst hij ook af, omdat dit in zijn ogen leidt tot het zien van technologie als niet meer dan een transparante barrière tussen mens en wereld, een middel tot een doel. Door de constituerende rol van technologie, wordt hoe mensen de wereld om zich heen waarnemen en de doelen die zij willen bereiken, echter medebepaald door de aanwezigheid van bepaalde technologieën of gereedschappen ('tools').²⁴

²¹ Peter-Paul Verbeek, 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research'. Gepubliceerd in: J.K. Berg O. Friis and Robert C. Crease, *Technoscience and Postphenomenology: The Manhattan Papers* (London 2016), pp. 189-204. Geraadpleegd op: <https://research.utwente.nl/en/publications/toward-a-theory-of-technological-mediation-a-program-for-postphen>, aldaar p.2-3.

²² Kiran, A., & Verbeek, P., 'Trusting Our Selves to Technology', in: *Knowledge, Technology & Policy*, 23:3 (2010), 409-427, aldaar 411.

²³ Kiran en Verbeek, 'Trusting Our Selves', 416-417.

²⁴ Ibidem, 412-416.

De mens heeft in deze situatie eveneens een rol. Voor Verbeek is het bestaan van technologie grotendeels een gegeven, iets waar weinig aan gedaan kan worden, maar de mens heeft wel invloed op hoe een bepaalde technologie gebruikt wordt, en hoe men kan omgaan met de technologische mediatie. Verbeek stelt echter dat het onmogelijk is om dat op een zinvolle manier te doen op basis van de aanname dat er een strikte scheiding is tussen 'mens' en 'technologie' als categorieën en het bestaan van een zekere tweestrijd tussen deze categorieën. Dit betekent echter niet dat technologie eenzijdig de menselijke belevingswereld bepaalt, er is wel degelijk ruimte voor een kritische blik op technologieën. Zoals Verbeek zelf schrijft: 'Mediatie kent geen uitweg, maar wel ruimte voor reflectie.'²⁵

De opvattingen van Verbeek over de onlosmakelijke verbondenheid van mens en technologie worden echter niet door iedereen gedeeld. Zo publiceerde Pieter Lemmens in 2017 een artikel waarin hij een neomarxistische kijk op de relatie tussen mens en technologie uiteenzet.²⁶ Lemmens beschouwt de huidige tijd als een periode van 'cognitief kapitalisme', waarin cognitieve capaciteiten worden ingezet om meerwaarde te genereren en waarin digitale netwerktechnologieën een grote rol spelen in het controleren en vastleggen van deze vormen van arbeid. Doordat deze technologieën van de gebruiker eisen dat deze continu verbonden is met het netwerk, wordt de aandacht van het subject continu gemobiliseerd en gefragmenteerd. Dit zorgt er vervolgens weer voor dat men de verbinding met het eigen en het sociale 'lichaam' kwijtraakt, waardoor solidariteit en collectief handelen tegengegaan worden.²⁷

Lemmens haalt wel enkele auteurs aan die wijzen op het mogelijk emancipatoire karakter van digitale netwerktechnologieën, maar voegt hier wel direct zijn eigen, uiterst kritische, noot aan toe. In tegenstelling tot Verbeek stelt Lemmens dat we technologie niet als iets onvermijdelijks moeten beschouwen, waar we een manier voor moeten vinden om ermee te leven, maar als iets dat aangepast moet worden aan de wil van mensen, als onderdeel van een revolutionair project tegen het hedendaagse kapitalisme.²⁸ In hun huidige staat, zo schrijft Lemmens, zijn digitale netwerktechnologieën namelijk slechts hulpmiddelen voor een toenemend repressief kapitalisme en werken ze individualisering en onverschilligheid in de hand. Verbeek schrijft in een reactie dat de nadruk die Lemmens legt op de grote, overkoepelende productiesystemen geen recht doet aan de ontwikkelingen op het gebied van techniekfilosofie in de afgelopen decennia, waarin analyse van de

²⁵ Peter-Paul Verbeek, *Op de vleugels van Icarus. Hoe techniek en moraal met elkaar meebewegen* (Rotterdam 2014) 11-14, citaat op p.14.

²⁶ Pieter Lemmens, 'Social Autonomy and Heteronomy in the Age of ICT: The Digital Pharmakon and the (Dis)Empowerment of the General Intellect', in : *Foundations of Science* 22 (2017) 287-296.

²⁷ Lemmens, 'Social Autonomy', 288-290.

²⁸ Ibidem, 294-295.

technologieën zelf juist centraal staat en waarin de strikte scheiding tussen mens en techniek waar Lemmens vanuit lijkt te gaan, wordt opgeheven.²⁹

Lemmens legt echter wel de vinger op een zere plek in het werk van Verbeek, namelijk dat er, door de nadruk van Verbeek op de constituerende rol van technologie en het beschouwen van technologie als een onvermijdelijkheid, er weinig ruimte lijkt te zijn voor menselijk handelen en menselijke agency. Op het eerste gezicht lijkt de theorie van Lemmens hier dan ook een goede uitkomst in te bieden: in plaats van te buigen voor technologische ontwikkelingen en ons eraan aan te passen, is het aan de mensheid om de regels op te stellen en technologie aan te passen zodat deze gebruikt kan worden voor het terugwinnen van de 'collectieve autonomie', zoals Lemmens het noemt.³⁰ Verbeek stelt hier echter tegenover dat een andere, aan Foucault ontleende, opvatting van het vrijheidsbegrip hier uitkomst biedt: vrijheid is geen afwezigheid van invloeden, maar een relatie tot deze invloeden, waarbij het subject zelf verantwoordelijk is voor het vormgeven van deze relatie.³¹

Zo is het in het werk van Verbeek dus wel degelijk mogelijk voor de mens om een actieve houding aan te nemen. Zijn ideeën zijn wellicht politiek wat minder scherp, of wat pragmatischer, dan de theorieën van Lemmens, maar zijn alles behalve bedoeld om de mens een ondergeschikte rol te geven, of menselijke agency weg te nemen. Hoewel een politieke analyse van het werk van Verbeek – door een marxistische lens – veel kan opleveren,³² lijken zijn critici niet te kunnen weerleggen dat er in het menselijk handelen altijd beïnvloedende factoren zijn en dat er situaties zijn waarin het zoeken naar een juiste verhouding tot die invloeden, het enige is dat men kan doen. Dit, gecombineerd met de nadruk op het analyseren van de technologieën zelf in het werk van Verbeek, maakt zijn theorie zeer geschikt als onderbouwing voor een praktische analyse als deze studie.

Machine learning en Big Data: de technologie en meer

Om, in lijn met Verbeek, de technologieën zelf als uitgangspunt te nemen, is het nodig om eerst de technologieën te specificeren. Dit kan, zeker in het geval van kunstmatige intelligentie, lastig zijn, aangezien de gebruikte terminologieën in het publieke debat en in de vakliteratuur soms nogal van elkaar kunnen verschillen. De vorm van kunstmatige intelligentie die in dit onderzoek besproken wordt, is data-analyse met behulp van *machine learning* algoritmes. Dit is nog steeds een wat brede

²⁹ Peter-Paul Verbeek, 'The Struggle for Technology: Towards a Realistic Political Theory of Technology', in *Foundation of Science* 22 (2017), 301-304.

³⁰ Lemmens, 'Social Autonomy', 295.

³¹ Verbeek, *Op de vleugels*, 108.

³² Bantwal Rao, M., Jongerden, J., Lemmens, P., & Ruivenkamp, G., 'Technological Mediation and Power: Postphenomenology, Critical Theory, and Autonomist Marxism', in: *Philosophy & Technology*, 28:3 (2015), 449–474.

omschrijving, maar in de bespreking van de casussen zal duidelijker worden welke praktische toepassingen hieronder vallen. Allereerst is het belangrijk om vast te stellen wat hier met 'algoritme' wordt bedoeld. In algemene zin is een algoritme niets meer dan een wiskundige procedure, een reeks instructies voor het verwerken van input. Het bijzondere van machine learning is echter dat een dergelijk algoritme zichzelf kan leren onderscheid te maken tussen een gewenst en ongewenst resultaat van deze verwerking, en daar gebruik van maakt om de eigen instructiereeks aan te passen.³³

Dit zelflerende karakter kan verregaande implicaties hebben voor de toepassing van dergelijke algoritmes in data-analyse en besluitvorming. Zo schrijft Lucas Introna over de problemen die dit oplevert voor het inzichtelijk maken van het algoritmisch proces: doordat de algoritmes veelal in samenwerkingsverband worden geschreven, of het werk later door andere programmeurs wordt overgenomen, en doordat ze de mogelijkheid hebben zichzelf aan te passen, wordt het heel moeilijk om het werk van de algoritmes te inspecteren en te begrijpen.³⁴ Introna stelt verder dat het hierdoor onmogelijk wordt om een eenduidige *agency* vast te stellen en spreekt van een 'agentic swarm' doordat zoveel verschillende partijen bij de ontwikkeling van het algoritme betrokken zijn. De broncode van hierdoor nooit als een eenvoudig gegeven worden beschouwd: '(...)source code is never simply the source of any action; rather, source code is only source code after the fact: its effectiveness depends on a whole imagined network of machines and humans.'³⁵

De omstandigheden die het creëren en de ontwikkeling van een algoritme omgeven, zijn dus vrij complex en lastig vast te stellen: het is geen eenduidig proces, maar vindt plaats in een uitgebreid netwerk van actoren. Dit kan ertoe leiden dat algoritmes snel ondoorzichtig worden. John Danaher schrijft over de mogelijke risico's die dit kan opleveren als dergelijke ondoorzichtige algoritmes (ook wel 'black boxes' genoemd) in een bestuurlijke context gebruikt worden.³⁶ Volgens Danaher leidt dit ertoe dat besluitvorming ondoorgroondelijk wordt en dat de mogelijkheden voor menselijk ingrijpen en burgerparticipatie sterk ingeperkt worden.³⁷ Met name dit laatste kan in zijn ogen desastreuze gevolgen hebben voor de legitimiteit van een democratisch systeem, waarvan participatie in, of in ieder geval inzichtelijkheid van, het besluitvormingsproces een belangrijke steunpilaar is.

Daarnaast is de wisselwerking tussen algoritmes en de sociaal-politieke context waarin ze gebruikt en ontwikkeld worden van groot belang. Dit blijkt uit een uitgebreide analyse van het

³³ Adam Greenfield, *Radical Technologies: The Design of Everyday Life* (New York 2017) 211-213.

³⁴ Introna, Lucas D., 'Algorithms, Governance, and Governmentality: On Governing Academic Writing', in: *Science, Technology & Human Values* 41 (2016), 17-49, aldaar 25.

³⁵ Introna, 'Algorithms', 24.

³⁶ John Danaher, 'The threat of algocracy: reality, resistance and accommodation' in: *Philosophy & Technology*, vol. 29 (2016), 245-268.

³⁷ Danaher, 'The threat', 249-252.

toenemend gebruik van algoritmische ‘anomaly detection’ in de veiligheidssector, in 2018 gepubliceerd door Claudia Aradau en Tobias Blanke.³⁸ Aradau en Blanke schrijven dat het detecteren van afwijkingen in grote hoeveelheden data de nieuwe grote trend in het denken over veiligheid is geworden, waarmee grote beloftes gepaard gaan van het kunnen detecteren van tot dan toe onbekende dreigingen, bijvoorbeeld op het gebied van terroristische activiteiten.³⁹ Zij leggen uit dat het zoeken naar deze ‘anomalies’ in werkelijkheid vooral neerkomt op het identificeren van uitschieters: datapunten of -patronen die relatief slecht te vergelijken zijn met de rest van de data. In plaats van de traditionele logica van moderne statistische analyse, die vooral uitgaat van een ‘normaal’ en afwijkingen daarvan, ontstaat er nu dus een nieuwe logica waarin vergelijkbaarheid de hoofdrol speelt.⁴⁰

Een voorbeeld hiervan dat Aradau en Blanke bespreken, is het gebruik van zogeheten clustering algoritmes. Dit zijn machine learning toepassingen, die data in clusters verdelen op basis van vergelijkbaarheid, gevisualiseerd in een grafiek, een zogeheten ‘feature space’. Hierbij leert het algoritme zelfstandig, zonder toezicht of van tevoren bedachte ideeën over de basis waarop de vergelijkbaarheid bepaald moet worden. Dit vraagt niet alleen om de verzameling van grote hoeveelheden data, en dus massasurveillance, maar ook om een vertrouwen in het vermogen van het algoritme om een wenselijk resultaat te produceren.⁴¹ De logica van vergelijkbaarheid levert echter lang niet altijd een goede uitkomst op. Dit is bijvoorbeeld te zien aan een zaak die in 2016 aan het licht kwam, toen bleek dat de Pakistaanse chef van Al-Jazeera op een Amerikaanse verdachtenlijst terecht was gekomen op basis van een algoritmische analyse van zijn sociale netwerken en reisgedrag, die in de ogen van het algoritme vergelijkbaar waren met mogelijke terroristen, zonder dat gekeken was naar de inhoudelijke context van zijn reizen en contacten. Dat de betreffende lijst dient als basis voor het vaststellen van doelen voor drone-aanvallen, laat zien hoe gevaarlijk het kan zijn om te vertrouwen op een dergelijke ‘afstandelijke’ analyse.⁴²

Een belangrijke onderliggende factor bij machine learning algoritmes, zijn de datasets waarmee het algoritme ‘getraind’ wordt. Door hun zelflerende karakter is het namelijk nodig om de algoritmes eerst een tijd trainingsdata te geven, zodat ze zich kunnen verfijnen en beter kunnen functioneren wanneer ze ‘echte’ analyses moeten maken. Hierbij komt een andere veelgebruikte term om de hoek kijken: de zogeheten ‘big data’. Deze term heeft misschien in letterlijke zin alleen betrekking op de toenemende omvang van beschikbare datasets, maar er is inmiddels veel meer

³⁸ Claudia Aradau & Tobias Blanke, ‘Governing others: Anomaly and the algorithmic subject of security’, in: *European Journal of International Security* 3:1 (2018), 1-21.

³⁹ Aradau en Blanke, ‘Governing others’, 7-8.

⁴⁰ Ibidem, 9-12.

⁴¹ Ibidem, 13-14.

⁴² Ibidem, 17-19.

betekenis omheen gekomen. Volgens danah boyd en Kate Crawford gaat het gebruik van de term vooral ook over ' (...)a capacity to search, aggregate, and cross-reference large data sets.'⁴³ Zij onderscheiden dan ook drie pijlers onder big data: technologie, analyse en mythologie, oftewel het geloof dat grotere datasets nieuwe, betere en objectievere kennis kunnen opleveren. Dit is met name belangrijk in combinatie met machine learning, aangezien machine learning algoritmes veelal gebruikt worden in het analyseren van 'big data'-sets en, andersom, datasets nodig zijn om de algoritmes te trainen.

Zoals boyd en Crawford al benoemen, is er sprake van een sterk geloof in de universele objectiviteit van grote datasets: cijfers liegen niet, en hoe groter de dataset, hoe groter de precisie van de analyse, ongeacht het onderwerp. Met big data wordt alles kwantificeerbaar en met kwantitatieve analyses is het aanzienlijk makkelijker om objectiviteit te claimen.⁴⁴ De beschikbaarheid van een dataset kan echter niet zomaar als een gegeven beschouwd worden: er zijn beslissingen genomen over het samenstellen van de dataset en veelal worden data aangepast, onder de noemer 'data cleaning', om onjuistheden of oneffenheden eruit te halen en het werk met de dataset te vergemakkelijken. Daarnaast heeft de dataset een herkomst, die vaak buiten beschouwing wordt gelaten, maar wat wel degelijk relevant is. boyd en Crawford gebruiken als voorbeeld Twitter, aangezien veel onderzoekers graag gebruik maken van de grote hoeveelheden tweets waartoe zij toegang kunnen krijgen. Twittergebruikers zijn echter geen doorsnede van de wereldbevolking en evenmin is de door Twitter geleverde set berichten een doorsnede van alles wat op Twitter staat: ook hierin is een geautomatiseerde selectie geweest.⁴⁵

Tyler Reigeluth trekt dit verder door en pleit voor het gebruik van de term 'digital trace' in plaats van 'data'.⁴⁶ Reigeluth pleit hier vooral voor omdat de term meer ruimte laat voor context en betekenis dan data, en omschrijft digital traces als 'fragments of past interactions or activities which, when correlated together, allow a preemption and prediction of future behaviors.'⁴⁷ Voor Reigeluth worden traces omgeven door materiële en technologische structuren en bestaat het dus nooit puur uit zichzelf.⁴⁸ Door de 'data' (of: traces) op deze manier in hun contexten te plaatsen, laat Reigeluth zien dat ze onmogelijk als gegeven beschouwd kunnen worden en nooit op zichzelf staan. Dit is ook van groot belang voor het gebruik van datasets bij machine learning, aangezien er daarbij wel vaak

⁴³ danah boyd & Kate Crawford, 'Critical Questions for BigData', in: *Information, Communication & Society* 15:5 (2012), 662-679, aldaar 663.

⁴⁴ Boyd en Crawford, 'Critical Questions', 666-667.

⁴⁵ Ibidem, 670-671.

⁴⁶ Tyler Reigeluth, 'Why data is not enough: digital traces as control of self and self-control', in: *Surveillance & Society* 12:2 (2014) 243-254, aldaar 248.

⁴⁷ Reigeluth, 'Why data is not enough', 250.

⁴⁸ Ibidem, 253.

naar data wordt gekeken alsof ze op zichzelf staan – het voorbeeld van de Al Jazeera chef spreekt hierbij boekdelen.

Het algoritme en het archief

Zoals Herbjørn Andresen schrijft, brengen moderne algoritmen ook nieuwe archiefwetenschappelijke uitdagingen met zich mee.⁴⁹ Andresen ziet een belangrijke rol voor archivering en informatiebeheer (of: *records management*): deze is niet zozeer om een uitleg te bieden van wat algoritmen doen, maar vooral om aan te wijzen in welke stukken en processen (delen van) een dergelijke uitleg te vinden zouden kunnen zijn.⁵⁰ Andresen erkent verder, evenals de bovenstaande auteurs, het belang van de context waarin het algoritme zich bevindt, ook wel de ‘data environment’ genoemd. Zeker waar het gaat om verantwoording, moet het volgens hem gaan om verantwoording voor het hele systeem, niet slechts een enkele component, bijvoorbeeld het algoritme.⁵¹

In een Nederlandse context is rond dit onderwerp enkele jaren geleden het nodige gepubliceerd in de bundel *Archives in Liquid Times*. Zo schrijft Charles Jeurgens in zijn bijdrage aan de bundel dat er een belangrijke taak bij de archivistiek ligt rondom het werken met algoritmen, gezien de traditionele rol van bewakers van verantwoording die archivariissen zichzelf toekennen.⁵² Jeurgens stelt dat de opkomst van ‘ubiquitous computing’ ertoe leidt dat bestaande definities binnen de archiefwetenschap niet meer voldoen. Zo is het oude denken over ‘records’ (‘archiefbescheiden’ in Nederlands juridisch jargon) gebaseerd op geschreven documenten die een vaste vorm op papier hebben, terwijl het digitale domein veel meer dynamiek kent, waarin dezelfde vastgelegde data tegelijkertijd op verschillende plekken en manieren gebruikt kunnen worden.⁵³

Dit zorgt ervoor dat er breder gedacht moet worden over wat er bewaard moet blijven en wanneer er sprake is van een goed bewaakte verantwoording. Zoals Jeurgens schrijft, moet het gaan over ‘(...) what people, institutions and communities want to be able to reconstruct for purposes of business, evidence, accountability and memory.’⁵⁴ In het geval van algoritmen betekent dit zeker ook een bredere blik, buiten de in- en outputdata en de broncode, omdat de context en de structuren waarbinnen het algoritme gebruikt wordt vanuit het perspectief van verantwoording minstens zo belangrijk zijn.⁵⁵ Arnoud Glaudemans en Jacco Verburgt schrijven hier ook over in hun

⁴⁹ Herbjørn Andresen, ‘A discussion frame for explaining records that are based on algorithmic output’, in: *Records Management Journal* 30:2 (2020), <https://doi.org/10.1108/RMJ-04-2019-0019>.

⁵⁰ Andresen, ‘A discussion frame’, 13.

⁵¹ Ibidem, 7.

⁵² Charles Jeurgens, ‘Threats of the data-flood. An accountability perspective in the era of ubiquitous computing’, in : Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times* (‘s-Gravenhage 2017) 196-211, aldaar 204-5.

⁵³ Jeurgens, ‘Threats of the data-flood’, 205.

⁵⁴ Ibidem, 207.

⁵⁵ Ibidem.

artikel voor de bundel, waarin zij onder meer pleiten voor een sterke wending tot het beheersen van digitale en virtuele ‘plekken’ als het gaat om informatiebeheer.⁵⁶ Hiermee bedoelen zij dat er minder aandacht moet zijn voor controle over de fysieke kant van digitale gegevens (de bits en hun fysieke dragers) en op ‘document-like records’, maar dat de nadruk moet liggen op controle over de technische- en informatiearchitecturen waarbinnen gewerkt wordt.⁵⁷

Uit het bovenstaande blijkt wel, dat de besproken technologieën en hun gebruik alles behalve op zichzelf staan. Verbeek legt een sterke nadruk op het analyseren vanuit de technologieën en de menselijke ervaring zelf, en dit biedt ook zeker een waardevol perspectief. Daarbij moet echter niet worden vergeten dat het gebruik van deze technologieën plaatsvindt in specifieke sociale, culturele, politieke en economische contexten. Deze beïnvloeden de manier waarop de mens de technologie gebruikt, maar dit werkt ook de andere kant op. Zoals Aradau en Blanke laten zien, kan een technologische eigenschap sturend zijn in een politiek proces, zoals het vormen en invullen van veiligheidsbeleid. Verbeek besteedt ook veel aandacht aan dit dubbelzinnige karakter van technologie en beschrijft de relatie tussen mens en technologie dan ook niet voor niets als een van wisselwerking en sterke onderlinge verbondenheid.

Het belangrijkste element van het werk van Verbeek, zoals dat hierboven uiteengezet is, is het vervagen van de grens tussen mens en technologie. Voor Verbeek is technologie en de menselijke afhankelijkheid ervan een onmiskenbaar onderdeel van de *human condition* en is de ethische opdracht voor de mens niet of we wel of geen technologische beïnvloeding in onze levens willen, maar hoe we daaraan vorm willen geven. Op het gebied van machine learning algoritmes en big data is dit laatste duidelijk ook nodig. Het gebruik van deze technologieën, en daarmee ook de invloed die ze kunnen uitoefenen op het menselijk leven, is in de laatste jaren enorm toegenomen, maar vaak nog met onvoldoende kritische reflectie op dit gebruik.

Dit laatste is juist bij overheidsorganisaties een groot risico: onvoldoende kritische reflectie op algoritmegebruik kan al snel leiden tot ondoorzichtig of oneerlijk gebruik van deze technologie. Zoals Danaher al schrijft, ondermijnt dit het democratisch proces en kan dit de legitimiteit van overheden ernstige schade toebrengen. De overkoepelende conclusie van al deze publicaties, van abstracte filosofische ideeën tot concrete casussen, is dat technologieën niet los gezien kunnen en mogen worden van de menselijke context waarin ze zijn ontstaan en gebruikt worden. Ditzelfde geldt voor hedendaagse algoritmen, zeker wanneer de gebruiker een overheidsorganisatie is, die taken

⁵⁶ Arnoud Glaudemans en Jacco Verburgt, ‘The Archival Transition from Analogue to Digital: Revisiting Derrida and Flusser’, in: Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times* (‘s-Gravenhage 2017) 120-137, aldaar 132.

⁵⁷ Glaudemans en Verburgt, ‘The Archival Transition’, 132-133.

uitvoert die directe weerslag hebben op (individuele) burgers. Om te zorgen dat het overheidshandelen in het uitvoeren van deze taken na te gaan is, zijn archieven onmisbaar, dus zullen zij ook een rol moeten spelen in het borgen van transparantie, eerlijkheid, verantwoording en uitlegbaarheid bij algoritmegebruik door overheden. Als dit handelen niet na te gaan is, kan er immers ook niet gekeken worden naar de context van het algoritmegebruik en kan er geen morele discussie gevoerd worden over de wenselijkheid ervan.

Hoofdstuk 2: Regulerende en beleidskaders

Het is niet verwonderlijk dat er nog weinig directe regulering is op het gebied van AI-gebruik. De technologie wordt nog maar sinds enige jaren in brede zin en op een zichtbare manier toegepast, waardoor enigszins ‘onder de radar’ van wetgevers kon blijven. De ontwikkeling van algoritmen is zo snel gegaan en heet zodanig complexe vormen aangenomen, dat wetgevers dit ook niet konden bijhouden. De laatste jaren wordt er echter op meerdere plekken gewerkt aan ethische principes, richtlijnen en eerste aanzetten tot wetgeving. Hoewel een regelgevend kader over het algemeen dus nog in de maak is, is het wel zinvol om te kijken naar die eerste aanzetten, aangezien die een idee kunnen geven van wat er aan wetgeving zou kunnen komen. Daarnaast geven dergelijke stukken ook aan welke waarden belangrijk worden geacht voor toekomstig beleid, wat van pas kan komen bij interne discussies en bewustwording van deze waarden.

Dit hoofdstuk bespreek dan ook een aantal relevante kaders op het gebied van algoritmen en archivering, met als doel om een antwoord te geven op de vraag wat voor kaders er zijn rondom overheidsinformatiebeheer en algoritmen, en hoe transparantie, uitlegbaarheid, verantwoording en eerlijkheid hierin terugkomen. Dit geeft een beter beeld van de huidige situatie, zodat duidelijker te definiëren is waar archivering een bijdrage kan leveren aan het verbeteren van deze situatie. Allereerst worden enkele wetten besproken die betrekking hebben op de archivering en openbaarmaking van overheidsinformatie en privacy: de huidige Nederlandse Archiefwet uit 1995, de komende vernieuwing van de Archiefwet, de Wet Open Overheid (WOO) en de Algemene Verordening Gegevensbescherming (AVG). Gezamenlijk vormen deze wetten een belangrijk deel van het kader voor het informatiebeleid van Nederlandse overheidsorganisaties. Er zijn natuurlijk meer wetten die op dit gebied relevant zijn, maar wegens tijdgebrek is gekozen om de nadruk te leggen op de wetten die het meest direct betrekking hebben op informatiebeheer. De WOO is natuurlijk nog niet ingevoerd, momenteel is nog altijd de Wet Openbaarheid van Bestuur (WOB) van kracht. Er is echter voor gekozen om prioriteit te geven aan de WOO omdat, ook al gaat het om toekomstige wetgeving, de onderliggende principes al redelijk vastliggen en deze een groot effect lijken te gaan hebben op de openbaarheid van Nederlandse overheidsinformatie.

In het volgende stuk worden enkele rapporten van het Rathenau Instituut en de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) besproken. Deze rapporten gaan over kwesties rondom publieke waarden en nieuwe technologieën en zijn bedoeld als adviezen voor Nederlandse beleidsmakers. Zo geven de rapporten dus een goed beeld van de richting die het beleid op zou kunnen gaan. Deze rapporten waren zowel binnen Nederland als daarbuiten veelbesproken, dus zullen ongetwijfeld invloed hebben (gehad) op huidige en toekomstige regelgeving. Vervolgens worden enkele publicaties van de Europese Commissie behandeld, die pogen een Europees

waardenkader voor AI op te zetten. De invoering van de AVG als Europese wetgeving laat zien dat de EU zich wil laten gelden in het digitale domein, dus zal een Europees kader ongetwijfeld weerslag hebben op de Nederlandse situatie. Tot slot bespreekt dit hoofdstuk de richtlijnen die het Ministerie van Justitie en Veiligheid in 2019 publiceerde, die een eerste vorm van regelgeving voor algoritmen zijn in een Nederlandse context.

De Archiefwet 1995

Een van de belangrijkste wettelijke kaders om te bespreken in het kader van dit onderzoek, is de Archiefwet. In deze wet is immers vastgelegd aan welke voorwaarden het archiefbeheer van Nederlandse overheidsorganisaties moet voldoen en vormt dus automatisch ook een belangrijk kader voor de archivering van algoritmen. De huidige Nederlandse Archiefwet stamt uit 1995, maar er wordt gewerkt aan een vernieuwing van de wet. De Archiefwet 1995 diende ter vervanging van de Archiefwet 1962 en verkortte onder meer de termijn waarbinnen overgedragen moesten worden naar een archiefdienst fors, van vijftig naar twintig jaar, een trend die gevolgd lijkt te worden door de Archiefwet 2021. Hoewel de Archiefwet 1995 dus vermoedelijk niet lang meer van kracht zal zijn, is het toch de moeite waard om te kijken naar de intenties achter deze wet, door bestudering van de Memorie van Toelichting. Deze geeft een idee van de principes en waarden die van belang zijn voor Nederlands archiefbeleid, evenals de nodige context voor de vernieuwing van 2021.

AI vroeg in de Memorie gaat het over het belang van archieven, niet alleen als ‘geheugen van de overheid’, maar ook als een belangrijke bouwsteen voor goed bestuur en een essentiële voorwaarde voor democratische controle.⁵⁸ Minister d’Ancona schrijft: ‘Zonder goed geordende en toegankelijke archieven zijn bestuurlijke zorgvuldigheid en continuïteit loze begrippen en is een goede democratische controle op het bestuur vrijwel onmogelijk.’⁵⁹ De wetshervorming van 1995 was vooral gericht op het verbeteren van de openbaarheid van de archieven en de voorwaarden waaronder archiefdiensten konden functioneren.⁶⁰ Om dit eerste doel te bereiken, is onder meer de overbrengingstermijn verkort en is de expliciete eis opgenomen om archieven niet alleen in goede en geordende, maar ook in toegankelijke staat te bewaren.⁶¹

Een interessante kwestie die de memorie aansnijdt, is die van privaatrechtelijke organisaties die publieke taken vervullen. Deze hebben, in de woorden van de minister, ‘zich een zodanige plaats verworven in ons staatsbestel, dat zij als een «verlengstuk» van de overheid moeten worden

⁵⁸ Ministerie van Welzijn, Volksgezondheid en Cultuur, *Memorie van Toelichting Archiefwet 1995* (Den Haag 1995), te raadplegen op: <https://www.inspectie-oe.nl/publicaties/publicatie/2014/11/12/memorie-van-toelichting-archiefwet-1995>, 2.

⁵⁹ Ibidem.

⁶⁰ Ibidem, 3.

⁶¹ Ibidem, 3-4.

beschouwd, zij het met een mate van autonomie.⁶² Dit betekent dat ze niet buiten de Archiefwet gehouden mogen worden, omdat burgers anders geen zicht meer zouden hebben op een significant deel van het handelen van de overheid. Een andere wijziging waarop de opstellers van de nieuwe wet hebben getracht te anticiperen, is het toenemen van digitaal werken. Dit doet men echter voornamelijk door op een aantal plekken de formulering 'ongeacht hun vorm' toe te voegen, bijvoorbeeld waar het informatiedragers of reproducties betreft, waarmee eigenlijk gezegd wordt dat er geen wezenlijk verschil is tussen informatie die op papier of digitaal is vastgelegd.⁶³ Elders stelt de Memorie ook dat digitale informatiedragers niet wezenlijk anders behandeld hoeven te worden dan papier op het gebied van preservatie.⁶⁴

Hoewel hierop anno 2020 het nodige is aan te merken, laat het wel de intentie achter de vernieuwing van de Archiefwet in 1995 zien. Dit was duidelijk om gebreken in de bestaande wetgeving af te vangen en om te zorgen dat volwaardige democratische controle mogelijk bleef. Dit is niet alleen te zien aan de toevoeging van digitale informatie, maar onder meer ook aan de verkorting van de overbrengingstermijn, de toegankelijkheidseis en de toevoeging van privaatrechtelijke organisaties. Deze moeten er voor zorgen dat overheidsarchieven vollediger en beter worden bewaard, zodat belangrijke kwaliteitscriteria als transparantie en controleerbaarheid van de overheid geborgd kunnen worden, evenals bestuurlijke continuïteit.

Vernieuwde Archiefwet, WOO en AVG

Voor de nieuwe Archiefwet is op het moment van schrijven nog geen definitieve tekst vastgesteld. Eind 2019 werd een eerste concept beschikbaar gesteld voor internetconsultatie. Op basis van de reacties hierop, wordt momenteel aan een nieuwe versie gewerkt. Het is dus nog niet mogelijk om precies te analyseren wat er in de nieuwe Archiefwet staat, maar de intenties ervan zijn al wel uiteen gezet in een concepttoelichting. Een van de belangrijkste, en meest besproken, wijzigingen, is het veranderen van de term 'archiefbescheiden' in 'documenten'.⁶⁵ Volgens de toelichting is dit onder meer bedoeld om aansluiting op de komende WOO mogelijk te maken en ook om te benadrukken dat er geen sprake is van een apart proces van archivering, maar van een samenhang in de processen die een informatieobject doorloopt.⁶⁶ Deze opvatting is in de archiefwetenschap steeds gangbaarder, waarin het doorgaans *archiving-by-design* wordt genoemd.

⁶² Ibidem, 14.

⁶³ Ibidem, 15.

⁶⁴ Ibidem, 4.

⁶⁵ Ministerie van Onderwijs, Cultuur en Wetenschap, *Van Archiefwet 1995 naar Archiefwet 202x: de belangrijkste wijzigingen van de nieuwe Archiefwet 202x, met korte toelichting*. (Den Haag 2019), te raadplegen op: <https://www.internetconsultatie.nl/archiefwet/document/5165>, 1-2.

⁶⁶ Ibidem, 2.

Hiernaast vallen twee verdere zaken op. Ten eerste verkort de nieuwe Archiefwet wederom de overbrengingstermijn, ten bate van 'de digitale duurzaamheid, de toegankelijkheid van de overheidsinformatie, de openbaarheid en het cultureel ergoed (sic).'⁶⁷ Er zijn echter mogelijkheden om de overdracht tweemaal tien jaar uit te stellen, als het gaat om stukken die nog regelmatig door de archiefvormer worden gebruikt. Ten tweede valt het op dat er een specifiek artikel in de nieuwe wet is opgenomen om de waarde van documenten als 'verantwoordingsinstrument en geheugen van de overheid' te benadrukken.⁶⁸ Er worden dus in de nieuwe wet zeker stappen gezet om met name digitale archivering en informatiebeheer meer onder de aandacht te brengen en het belang hiervan verder duidelijk te maken. Grote stappen als het stimuleren (of verplichten) van *archiving-by-design* en duurzame toegankelijkheid, ontbreken echter. Algoritmen en data worden ook niet behandeld in het concept, wat zeker een gemiste kans is, aangezien er al enige jaren gepraat wordt over algoritmegebruik door overheden.

In de concepttoelichting gaat het ook geregeld over de aansluiting met andere wetten, zoals de WOO en de AVG. Er is hier geen ruimte om deze volledig te analyseren, dus zal dit stuk zich beperken tot een bespreking van de achterliggende ideeën en waarden van deze wetten. In de meest recente memorie van toelichting bij de WOO wordt als eerste reden genoemd dat Nederland achterop raakt op het gebied van open overheidsinformatie en dat er onvoldoende actief beschikbaar wordt gesteld.⁶⁹ De WOO verplicht dan ook tot grootschalige actieve openbaarmaking, in machineleesbare formaten ten behoeve van hergebruik, en het bijhouden van een register van 'documenten en datasets' waarover overheden beschikken.⁷⁰ Hierin staan dus waarden als transparantie en openbaarheid centraal, met een opvallende toevoeging dat het publiek beschikbaar stellen van overheidsinformatie ook economische waarde kan hebben, door hergebruik hiervan door commerciële partijen.⁷¹ Algoritmen lijken nog geen speciale aandacht te krijgen in de WOO, maar de eis van het register van datasets zal zeker invloed hebben op de transparantie rondom algoritmegebruik.

Wat betreft de AVG zijn de resultaten waarschijnlijk niet erg verrassend. De wet, vrijwel universeel aangeduid als 'de privacywet', heeft als grote doel gegevensbescherming op individueel niveau, of, in de woorden van een groep vooraanstaande informatiejuristen, 'to make the electronic body inviolable.'⁷² Principes als dataminimalisatie en doelgebondenheid van dataverzameling

⁶⁷ Ibidem, 7.

⁶⁸ Ibidem, 2.

⁶⁹ *Kamerstukken II 2013/14*, 33328, 3.

⁷⁰ Ibidem, 5.

⁷¹ Ibidem, 53.

⁷² C.J. Hoofnagle, B. van der Sloot, & F. Zuiderveen Borgesius, 'The European Union General Data Protection Regulation: What It Is And What It Means', in: *Information & Communications Technology Law*, 28:1 (2019), 65-98, aldaar 67.

moeten ervoor zorgen dat alleen voor een specifiek doel noodzakelijke informatie over individuen verzameld wordt, en dat deze op een eerlijke en rechtmatige wijze verwerkt wordt.⁷³ Verder vraagt de AVG ook om transparantie over dataverzameling en -verwerking: er is expliciete toestemming voor nodig en op verzoek van een individu moet een datacontroller tijdig een overzicht kunnen aanleveren van de gegevens die de controller van dat individu heeft. Deze vormen van transparantie moeten bovendien ook actief zijn, niet alleen wanneer het individu erom vraagt.⁷⁴

Adviezen voor verantwoord technologiegebruik binnen Nederland

In Nederland wordt al enige jaren door diverse adviesorganen gepubliceerd over de maatschappelijke uitdagingen die het gebruiken van nieuwe technologieën met zich meebrengt. Een belangrijke publicatie op dit gebied, was het rapport *Big Data in een vrije en veilig samenleving* van de WRR.⁷⁵ De auteurs van dit rapport schrijven hierin uitgebreid over de mogelijkheden die Big Data toepassingen bieden in het veiligheidsdomein, evenals de risico's die gepaard gaan met het implementeren van deze technologieën voor grootschalige dataverzameling en -analyse. Een belangrijk risico dat zij noemen, is het ontstaan van een zogeheten transparantieparadox: burgers worden, door middel van grootschaligere dataverzameling en -analyse, steeds transparanter voor overheden, terwijl de gebruikte technieken voor verzameling en analyse juist minder transparant worden.⁷⁶ Hiermee doelen de auteurs ook specifiek op algoritmen, die met name in het veiligheidsdomein vaak geheimgehouden worden. Verder benoemden ze expliciet de noodzaak om een mens in het proces betrokken te houden.⁷⁷

Het rapport *Opwaarderen* van het Rathenau Instituut gaat in bredere zin over de maatschappelijke effecten van nieuwe, digitale technologieën en over de zogeheten 'blinde vlekken' die er op dit gebied zijn in de Nederlandse *governance*-structuur.⁷⁸ De auteurs doen op basis van hun analyse weinig concrete aanbevelingen voor hoe de risico's die hierdoor ontstaan het hoofd geboden moeten worden, maar roepen vooral op tot brede politieke en maatschappelijke discussies over de gevolgen van 'digitalisering' (zoals zij het geheel aan besproken technieken noemen) voor belangrijke publieke waarden.⁷⁹ Bij deze discussies zouden vooral ook bedrijven en maatschappelijke

⁷³ Hoofnagle e.a., 'The European Union', 77-78.

⁷⁴ Ibidem, 86.

⁷⁵ Wetenschappelijke Raad voor het Regeringsbeleid, *Big Data in een vrije en veilige samenleving* (Amsterdam 2017).

⁷⁶ WRR, *Big Data*, 93.

⁷⁷ Ibidem, 71-72.

⁷⁸ Kool, L., J. Timmer, L. Royakkers en R. van Est, *Opwaarderen- Borgen van publieke waarden in de digitale samenleving*. (Den Haag, Rathenau Instituut 2017).

⁷⁹ Kool, Timmer e.a., *Opwaarderen* 133.

organisaties betrokken moeten zijn, en ze zouden gepaard moeten gaan met een verbetering van het toezicht op het digitale domein.⁸⁰

De aanbevelingen zijn redelijk algemeen en vooral gericht op de maatschappij in brede zin, maar zijn vooral zo geformuleerd omdat de problematiek waarop de auteurs ingaan, ook erg breed voorkomt. In een eerdere publicatie schrijft het Rathenau over het behoud van mensenrechten in ‘the Robot Age’, waarbij ook aandacht is voor de rol van algoritmen.⁸¹ De auteurs spreken vooral hun zorgen uit over de mogelijkheid om mensen automatisch in categorieën of groepen in te delen met behulp van algoritmen, samen met de ondoorzichtigheid waarmee dit proces momenteel vaak plaatsvindt, zoals bij veel internetbedrijven.⁸² Hierdoor weten individuen vaak überhaupt niet dat ze op deze manier geprofileerd worden, laat staan hoe, en is het ook lastig om aan te tonen dat er sprake is van onrechtmatige discriminatie, aldus het rapport. De auteurs pleiten voor het bevorderen van wat zij *algorithmic accountability* noemen, waarvoor in hun ogen ‘betekenisvolle transparantie vereist is.’⁸³ De auteurs gaan helaas niet verder in op wat ze hiermee bedoelen, maar evenals het *Opwaarderen*-rapport geeft deze analyse wel een aanzet voor verdere ideeën en debatten over welke invulling van deze concepten wenselijk zou zijn, maar verantwoording en eerlijkheid komen wel duidelijk terug in hun argumentatie.

Human-centric AI: een Europees raamwerk

Op Europees niveau wordt ook veel over AI nagedacht. De Europese Commissie publiceert sinds enkele jaren geregeld stukken over de rol van AI voor de toekomst van de Europese Unie (EU), waarin ook een aanzet wordt gemaakt tot het vormgeven van een passende ‘Europese’ rol voor AI in de samenleving. De meest recente uitkomst hiervan is een white paper, in februari 2020 gepubliceerd.⁸⁴ Dit stuk zet een Europese benadering van AI uiteen, waarin investering en regulering op Europees niveau ervoor moeten zorgen dat de EU een wereldwijd leidende positie inneemt op het gebied van kunstmatige intelligentie en wat de auteurs de ‘dataeconomie’ noemen.⁸⁵ Het

⁸⁰ Ibidem.

⁸¹ Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool, *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality—Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)* (Den Haag 2017).

⁸² Van Est en Gerritsen, *Human rights*, 40-41.

⁸³ Ibidem, 41.

⁸⁴ Europese Commissie, *White Paper On Artificial Intelligence – A European approach to excellence and trust* (Brussel, 19 februari 2020), 1-2; 10. Te raadplegen op: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

⁸⁵ Europese Commissie, *White Paper*, 1-2.

uitgangspunt voor de regulering moet zijn dat EU-burgers in AI kunnen vertrouwen, door AI zodanig te implementeren dat de mens centraal staat.⁸⁶

Deze benadering, ook wel ‘human-centric AI’ genoemd, is in een eerder stuk reeds verder uitgewerkt.⁸⁷ Hierin staan enkele voorwaarden die nodig zouden zijn om AI *human-centric* te maken en om te zorgen dat er vertrouwen hierin gecreëerd wordt. Naast eisen op het gebied van zaken als eerlijkheid, non-discriminatie en menselijk toezicht, gaat het stuk ook verder in op privacy, databeheer en transparantie. Op het gebied van databeheer gaat het deels om kwaliteit en integriteit, maar ook om de mogelijkheid van individuele burgers om volledige controle over hun data te hebben.⁸⁸ Wat betreft transparantie spreken de auteurs van maximale uitlegbaarheid van een AI-systeem aan alle betrokken partijen – inclusief burgers die met het systeem te maken zouden kunnen krijgen.⁸⁹

De expertgroep AI van de Europese Commissie heeft nog verdere richtlijnen gepubliceerd waaraan AI-systemen zouden moeten voldoen om betrouwbaar (‘trustworthy’) te zijn.⁹⁰ Aan de inmiddels bekende eis van transparantie, voegen zij hier ook expliciet de eis van traceerbaarheid toe. In de woorden van de auteurs: ‘The data sets and the processes that yield the AI system’s decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency.’⁹¹ Hierin komt dus een duidelijke voorwaarde op het gebied van archivering naar voren. Helaas noemen de auteurs niet wat hier de ‘best possible standard’ zou kunnen zijn, maar het is wel duidelijk dat het voor hen een belangrijk onderdeel is van het toewerken naar een betrouwbare implementatie van AI. Hiermee missen de auteurs wel een kans om een koppeling te maken met bestaande standaarden, zoals de ISO 15489 standaard voor records management.⁹²

Dit komt eveneens terug in de white paper, waarin aparte aandacht wordt besteed aan de eisen voor ‘keeping of records and data’.⁹³ De auteurs bevestigen hierbij het belang van goede archivering, zodat problematische handelingen of beslissingen door AI-systemen getraceerd kunnen worden. Het stuk noemt daarnaast enige categorieën die bewaard zouden moeten worden. Ten eerste zouden er ‘accurate records’ over de trainingsdataset bewaard moeten worden, inclusief een

⁸⁶ Ibidem, 3-4.

⁸⁷ Europese Commissie, *Building Trust in Human-Centric Artificial Intelligence* (Brussel, 8 april 2019). Te raadplegen op: <https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-168-F1-EN-MAIN-PART-1.PDF>.

⁸⁸ Europese Commissie, *Building Trust*, 5.

⁸⁹ Ibidem, 5.

⁹⁰ High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (Brussel 2019). Te raadplegen op: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

⁹¹ High-Level Expert Group, *Ethics Guidelines*, 20.

⁹² Zie: <https://www.iso.org/standard/62542.html>.

⁹³ Europese Commissie, *White Paper*, 19-20.

beschrijving van de dataset en hoe deze tot stand gekomen en geselecteerd is. Ten tweede zouden, in uitzonderlijke gevallen, de datasets zelf bewaard moeten worden. Tot slot zouden de methodologieën die gebruikt zijn om het algoritme te programmeren en te trainen – waaronder ook testen en valideren vallen – gedocumenteerd moeten worden.⁹⁴ Binnen het Europese raamwerk-in-aanbouw wordt dus ook erkend dat voor AI-systemen die door burgers vertrouwd kunnen worden, goede en betrouwbare documentatie vereist is, zodat de werking van de systemen uitlegbaar en traceerbaar is, maar wederom wordt er niet gesproken over mogelijke standaarden op het gebied van archivering.

De richtlijnen van het Ministerie van Justitie en Veiligheid

In Nederland heeft het Ministerie van Justitie en Veiligheid (J&V) in 2019 een eerste aanzet gegeven tot richtlijnen voor het gebruik van algoritmen door Nederlandse overheidsorganisaties, als reactie op het eerder besproken WRR-rapport uit 2016.⁹⁵ Een van de eerste zaken die in de richtlijnen benoemd worden, is het belang van het begrip ‘uitlegbaarheid’. De richtlijnen contrasteren dit begrip met technische transparantie, waarmee bijvoorbeeld het publiceren van de broncode wordt bedoeld. In tegenstelling tot technische transparantie, vraagt uitlegbaarheid om uitgebreidere en gedetailleerdere openheid, zodat ‘(...)de inzet, het doel en de uitkomst van een algoritme in begrijpelijke taal (...)’ uitgelegd kunnen worden.⁹⁶ Dit is vooral belangrijk bij algoritmen die gevolgen kunnen hebben voor individuele burgers en plaatst een ogenschijnlijk harde bovengrens op de complexiteit van algoritmen die gebruikt kunnen worden: zodra het te complex is om uit te leggen, mag het niet meer gebruikt worden.

Bij een verdere toelichting wordt onder meer genoemd dat hier ook bij hoort dat gehanteerde aannames vastgelegd worden. De keuze voor het principe wordt ook verder onderbouwd: volgens de richtlijnen, zijn overheden verantwoordelijk voor de uitkomsten van de door hen gebruikte algoritmen, ook als het gaat om algoritmen die door een andere partij gemaakt worden, wat betekent dat de algoritmen gerechtelijk getoetst moeten kunnen worden. Zonder voldoende documentatie, is dat niet mogelijk.⁹⁷ Verderop wordt nogmaals nadrukkelijk benoemd: ‘het algoritme dient gedocumenteerd te zijn’, zodat het toegankelijk is voor ‘toezichthouders, experts en burgers’.⁹⁸ De selectie van de brondata hoort hier ook bij: er mag alleen relevante inputdata

⁹⁴ Ibidem, 19-20.

⁹⁵ Ministerie van Justitie en Veiligheid, *Bijlage bij Brief over waarborgen tegen risico's van data-analyses door de overheid* (Den Haag 2019). Te raadplegen op: <https://www.rijksoverheid.nl/documenten/rapporten/2019/10/08/tk-bijlage-over-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid>.

⁹⁶ Ministerie van Justitie en Veiligheid, *Bijlage bij Brief over waarborgen*, 2-5, citaat op p.3.

⁹⁷ Ibidem, 6.

⁹⁸ Ibidem, 11.

gebruikt worden, en de keuze en onderbouwing hiervan dient gedocumenteerd te worden.⁹⁹

Wanneer het gaat over exacte procedures rondom het opstellen van deze documentatie en de omgang met commerciële partners, die een belangrijke rol spelen in deze kwestie, blijven de richtlijnen echter nog wel in gebreke, ook als het gaat om archivering.

Verder staan er andere principes in de richtlijnen, die goed aansluiten op het idee van uitlegbaarheid. Zo staat er bij het principe van auditeerbaarheid genoemd dat ‘Modellen, algoritmes, data en beslissingen met specifieke consequenties voor individuele burgers moeten worden vastgelegd,’ waar archivering wederom belangrijk voor is.¹⁰⁰ De richtlijnen benoemen verder dat het nodig is om gebruikte modellen en algoritmen te valideren voor ze in gebruik worden genomen en ze daarna regelmatig te blijven testen. Tot slot benadrukken de richtlijnen dat de overheid het verplicht is om voldoende transparant te zijn over de data-analyses die ze uitvoert. Hierbij is het belangrijk dat de informatie duidelijk vindbaar en in heldere, begrijpelijke taal geformuleerd is. De voorlichting hierover zou, volgens de richtlijnen, via de website van de betreffende overheidsdienst moeten plaatsvinden.¹⁰¹

Al met al bieden de richtlijnen een goede aanzet voor verdere regulering van algoritmegebruik door overheidsorganisaties. Zoals een recente impactanalyse door VNG Realisatie echter benoemt, gaan de richtlijnen nog teveel over wat er moeten gebeuren, en te weinig over hoe dat precies moet gebeuren, of wanneer er sprake is van voldoende maatregelen.¹⁰² Dit is terechte kritiek: de richtlijnen stellen regelmatig eisen die verder weinig gekwalificeerd worden, zoals: ‘zorg voor reproduceerbaarheid’, waarbij verder niet wordt toegelicht wat reproduceerbaarheid in deze context inhoudt.¹⁰³ De vorm van transparantie die de richtlijnen voorstellen, is ook zeker niet zonder beperkingen. Zoals Mike Ananny en Kate Crawford schrijven, kent transparantie ook de nodige beperkingen, zeker wanneer het gaat om geavanceerde algoritmen.¹⁰⁴ Hierbij is, zo schrijven zij, ook interactie met het systeem nodig om het te leren kennen en er echt inzicht in te krijgen. Technische factoren kunnen dit nog verder bemoeilijken, bijvoorbeeld bij *deep learning*.

Het belang van inzichtelijkheid van algoritmen is nergens duidelijker geworden dan in de geruchtmakende rechtszaak rondom het Systeem Risico Indicatie (SyRI) dat gebruikt werd door Nederlandse overheden voor fraudedetectie in het sociaal domein. Dit werd door de rechtbank onrechtmatig verklaard, onder meer door de mate van gegevensuitwisseling die plaatsvond, maar

⁹⁹ Ibidem, 10.

¹⁰⁰ Ibidem, 6.

¹⁰¹ Ibidem, 11-14.

¹⁰² VNG Realisatie, *Impactanalyse Richtlijnen voor het toepassen van algoritmes door gemeenten* (Den Haag 2020) 3-4.

¹⁰³ Ministerie van Justitie en Veiligheid, *Bijlage bij brief over waarborgen* 11.

¹⁰⁴ M. Ananny & K. Crawford, ‘Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability’, in: *New Media & Society* 20:2 (2018), 973–989, aldaar 981-982.

ook vanwege het gebrek aan openheid in het project: de rechtbank stelde in haar vonnis dat de precieze werking van SyRI niet te toetsen was door de sterke mate van geheimhouding rondom de gebruikte modellen.¹⁰⁵ Enige tijd geleden doken vergelijkbare bezwaren op tegen een soortgelijk systeem dat gebruikt wordt in de gemeente Nissewaard.¹⁰⁶ Dit algoritme is ontwikkeld door een commerciële partij en wordt eveneens gebruikt voor fraudedetectie in het sociaal domein. Het algoritme is niet inzichtelijk en de beschikbare technische documentatie is te complex om van wezenlijke uitlegbaarheid te kunnen spreken.¹⁰⁷ Waar SyRI leidde tot een rechtszaak, blijft het in het geval van Nissewaard nog bij protesten in de gemeenteraad, maar de klacht is hetzelfde: grootschalige dataverwerking door een *black box* algoritme met mogelijk significante gevolgen voor burgers.

Uit de bovenstaande analyse blijkt wel dat de vier criteria goed verankerd zijn, of ten minste verbonden zijn, met bestaande wet- en regelgeving. Bij de Archiefwet speelt verantwoording een hoofdrol. De geplande veranderingen wijzen erop dat men hieraan blijft werken, bijvoorbeeld met de verkorting van de overbrengingstermijn. Bij de AVG en in het bijzonder de WOO komt transparantie meer naar voren, verbonden met het idee van een open en controleerbare overheid. De rapporten van het Rathenau en de WRR wijzen goed op de gevaren die gebrekkige regelgeving op dit gebied met zich mee zou kunnen brengen. Zij zien de oplossingen vooral in toenemende transparantie van de overheid, het voeren van een maatschappelijk debat over de plek die nieuwe technologieën zouden moeten krijgen in de samenleving en het zorgen voor voldoende controlebaarheid en traceerbaarheid van algoritmen, zodat er verantwoording voor hun gebruik kan worden afgelegd. Dit komt ook terug in de stukken van de Europese Commissie, die met het idee van *human-centric AI* criteria als transparantie, eerlijkheid en controlebaarheid samenbrengen, evenals grondrechten als privacy en het recht op non-discriminatie. Zij benadrukken bovendien het belang van goede documentatie voor het in de praktijk brengen van deze criteria.

De opvallendste toevoeging van de richtlijnen van J&V is de eis dat algoritmen uitlegbaar moeten zijn. Dit heeft deels te maken met de forse impact die deze eis zou kunnen hebben, maar ook doordat er weinig concreets in de richtlijnen te vinden is over hoe deze uitlegbaarheid in de praktijk

¹⁰⁵ Rechtbank Den Haag, uitspraak ECLI:NL:RBDHA:2020:865 (5 februari 2020): <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865>. Zie onder meer artikel 6.49 en 6.65.

¹⁰⁶ Adriaan de Jonge, “Rookgordijn’ rondom fraudesysteem Nissewaard”, in : *Binnenlands Bestuur*, 24 april 2020. Te raadplegen op: <https://www.binnenlandsbestuur.nl/sociaal/nieuws/rookgordijn-rondom-fraudesysteem-nissewaard.13026523.lynkx>. Laatst geraadpleegd op 27 juli 2020.

¹⁰⁷ Totta Data Lab, *Technische documentatie voorspelmodel bijstandsfraude Gemeente Nissewaard* (2019). Te raadplegen op: https://nissewaard.notubiz.nl/document/8567516/1/Beantwoording_LOB%2C_SyRI-uitspraak_voor_algoritme_Totta_Data_Lab%2C_bijlage. Laatst geraadpleegd op 27 juli 2020.

moet worden gebracht. Het laat echter wel zien dat er een duidelijk kader van wetten, onderzoeksrapporten en richtlijnen zijn die openheid, transparantie, eerlijkheid en controleerbaarheid van overheden ondersteunen. In het geval van de richtlijnen van J&V gaat het nog verder dan alleen bestaande informatie openbaar maken, maar ook om het actief zorgen dat gebruikte algoritmen uitlegbaar zijn. In al deze kaders speelt goede documentatie bovendien een belangrijke rol, als voorwaarde voor het voldoende borgen van deze kwaliteitscriteria.

Helaas ontbreekt het nog te vaak aan precieze uitwerkingen van de kwalitatieve eisen waar digitale overheidsinformatie aan zou moeten voldoen. Er wordt wel geregeld gesproken over de behoefte aan 'goede documentatie', maar wat dat inhoudt blijft onduidelijk, zowel op Nederlands als op Europees niveau. Verder wordt er weinig gezegd over wat er met de gemaakte documentatie moet gebeuren, oftewel het proces van archivering. Dat is een groot gemis in deze kaders en zorgt ervoor dat er in de praktijk veel aan de interpretatie overgelaten wordt. Dit is met name te zien bij de richtlijnen van J&V, die met uitlegbaarheid een belangrijk en potentieel zwaar ingrijpend concept introduceren, maar vervolgens weinig toelichten over hoe dit in de praktijk zou moeten worden gebracht. Bij de vernieuwing van de Archiefwet valt vooral op dat er nog op weinig manieren aandacht wordt besteed aan de manieren waarop de overgang naar volledig digitaal werken de omgang met informatie beïnvloedt. Er lijkt een aanname te zijn dat de principes voor digital en papier grotendeels hetzelfde kunnen blijven, terwijl uit hoofdstuk 1 duidelijk blijkt dat dit niet het geval is. Zolang deze gebreken blijven bestaan, zullen er gaten blijven in het wettelijk kader voor informatiebeheer en de omgang met algoritmen.

Hoofdstuk 3: Interviews

Dit hoofdstuk behandelt de resultaten van de gehouden interviews. De interviews hadden een tweeledig doel: enerzijds informatie verzamelen van ambtenaren die ervaring hebben met het werken in projecten waarbij algoritmen worden gebruikt, en anderzijds het verzamelen van informatie over hoe archivering binnen Nederlandse gemeenten in het algemeen plaatsvindt, vooral met het oog op het digitale domein. Nederlandse gemeenten hebben allemaal op wisselende wijze in de afgelopen tien jaar de overgang gemaakt naar digitaal werken, wat betekent dat er dus ook digitale archieven gevormd moeten worden. Deze digitalisering zorgt echter zowel binnen de archiefwereld als daarbuiten voor discussies en onverwachte obstakels. Daarom was het voor dit onderzoek ook belangrijk om dit in kaart te brengen, om te zien waar de adviezen op het gebied van algoritmen moesten aansluiten. In dit hoofdstuk worden de interviews geanalyseerd, met als doel om de vragen te beantwoorden welke problemen er in de dagelijkse praktijk zijn op het gebied van archivering en informatiebeheer binnen gemeenten en hoe er momenteel met algoritmen wordt gewerkt binnen gemeenten.

Dit hoofdstuk bespreekt de belangrijkste bevindingen uit de interviews. Het eerste deel van het hoofdstuk bespreekt archivering en informatiebeheer bij gemeenten. Dit stuk is voor een belangrijk deel gebaseerd op de gesprekken met archief- en informatieprofessionals, die veel konden vertellen over hoe deze taken in algemene zin verlopen bij gemeenten. Daarnaast waren de gesprekken met medewerkers van algoritmeprojecten hierin ook erg nuttig, aangezien deze veel konden vertellen over de dagelijkse praktijk van informatiebeheer bij mensen die andere primaire taken hebben. In het tweede deel komen zij nog meer aan het woord, aangezien hier gesproken wordt over diverse aspecten van algoritmeprojecten bij gemeenten. Het gaat hierbij veel over het doel van en de gedachte achter deze projecten, maar ook over andere belangrijke thema's als de invulling van het begrip 'uitlegbaarheid' en samenwerking met commerciële partijen.

Methode

Er zijn grofweg twee 'typen' geïnterviewden te onderscheiden: archief- en informatieprofessionals, werkzaam bij gemeenten of archiefdiensten,¹⁰⁸ en ambtenaren die betrokken zijn geweest bij projecten met algoritmen. In deze laatste groep liepen de functies behoorlijk uiteen, van managers voor het implementeren van datagedreven werken, tot IT-projectleiders en medewerkers van de

¹⁰⁸ Een archiefdienst is een instelling waaraan overheidsarchieven overgedragen moeten worden en die de verplichting heeft deze archieven te beheren en beschikbaar te houden. Zie ook de omschrijvingen hier: <https://www.archieven.nl/nl/help?id=24>; <https://archiefwiki.pleio.nl/wiki/Archiefdienst>. De Archiefwet 1995 spreekt van 'archiefbewaarpplaats': <https://www.nationaalarchief.nl/archiveren/kennisbank/archiefbewaarpplaats>.

gemeentelijke afdeling Onderzoek en Statistiek. De geïnterviewden zijn op verschillende manieren geselecteerd. Bij de archiefprofessionals is de lijst tot stand gekomen met behulp van dhr. Jeurgens, die geholpen heeft met het selecteren van mensen met ervaring met archiveren op gemeentelijk niveau, die actief zijn binnen de beroepsgemeenschap en onder meer betrokken zijn bij werkgroepen waarin over toekomstbestendig archiefbeleid wordt nagedacht. Bij de betrokkenen bij algoritmeprojecten is de selectie gemaakt op basis van deelnemerslijsten van eerdere meewerksessies van het project Eerlijke Algoritmen. In totaal zijn er 13 mensen geïnterviewd: 7 uit het archief- en informatieveld en 6 uit de algoritmeprojecten. Voor de medewerkers van algoritmeprojecten was het vooral belangrijk om mensen in verschillende posities te spreken, om verschillende perspectieven op deze projecten te krijgen, van de strategische plannen die eromheen staan tot de dagelijkse werkzaamheden in de praktijk. Op het gebied van archivering en informatiebeheer zijn vooral hogere functionarissen geïnterviewd, omdat het hierbij belangrijker was een algemeen overzicht te krijgen van de stand van zaken in diverse gemeenten. Het was wel belangrijk om perspectieven uit zowel grote als kleinere gemeenten te krijgen, zodat eventuele verschillen hiertussen in kaart konden worden gebracht en om te voorkomen dat de resultaten teveel toegespitst zouden zijn op een bepaald type gemeente.

De interviews hadden een semigestructureerde vorm. Voor beide categorieën geïnterviewden was een lijst opgesteld met mogelijke vragen, discussiepunten en manieren om door te vragen, maar deze werd niet systematisch doorgewerkt.¹⁰⁹ De bedoeling van deze methode was om een grotendeels vrij gesprek te voeren, waarbij de geïnterviewden zoveel mogelijk ruimte kregen om zich uit te spreken en waar ook de mogelijkheid open bleef om op bepaalde onvoorbereide punten door te vragen.¹¹⁰ Hierbij kon de lijst (of: *interview guide*) dienen als hulpmiddel om het gesprek op gang te brengen en om te voorkomen dat er teveel van het onderwerp van het onderzoek afgeweken werd. De interviews vonden telefonisch plaats of in een videobelgesprek, aangezien een fysieke ontmoeting niet mogelijk was door de destijds geldende lockdownmaatregelen. Op basis van opnames en aantekeningen zijn de interviews uitgewerkt. Nu het onderzoek is afgerond, zijn deze vernietigd in het belang van de privacy van de geïnterviewden.

Archivering binnen gemeenten

Een van de thema's die bij de gesprekken met archiefprofessionals snel naar voren kwam, was de inkoop van nieuwe digitale applicaties en systemen. Meerdere geïnterviewden noemden dat zij zelden tot nooit geraadpleegd werden bij dit proces, tenzij ze er zelf het initiatief voor namen. Dit

¹⁰⁹ De vragenlijsten, of *interview guides*, zijn terug te vinden in de bijlage.

¹¹⁰ Alison Jane Pickard, *Research Methods in Information Science* (Londen 2013) 196.

heeft tot gevolg dat belangrijke functies voor archiefbeheer, zoals een goed werkende vernietigingsfunctie, vaak ontbreken.¹¹¹ Dit is problematisch, vooral omdat men ook ziet dat er steeds meer behoefte komt aan specifieke applicaties, gericht op bepaalde processen, waardoor centrale systemen als een zaakstelsel of Documentmanagementsysteem (DMS) minder gebruikt worden, terwijl dit juist de systemen zijn waarin deze functionaliteiten voor archiefbeheer het beste zijn verwerkt.¹¹² Een geïnterviewde noemde hiervoor als mogelijke verklaring dat deze systemen vertrouwen op handmatige invoering van de stukken en mogelijke metadata, waar velen geen tijd voor kunnen of willen maken.¹¹³

Dit wordt ook bevestigd door de geïnterviewden die niet in het archiefwezen werkzaam zijn. Stukken die naar een wethouder of naar de gemeenteraad gaan, worden in het zaakstelsel gezet, maar verder blijft alles binnen de procesapplicaties, netwerkschijven of samenwerkingstools zoals Sharepoint.¹¹⁴ Een geïnterviewde, werkzaam bij een afdeling onderzoek en statistiek, zei dat als er werk van de afdeling naar de raad ging, of onderdeel werd van een beleidsplan, de aanvullende handelingen die nodig waren om de stukken in het 'officiële' systeem te plaatsen, meestal door beleidsmedewerkers werden uitgevoerd, aangezien zij er beter in waren.¹¹⁵ Dit laat zien dat het documenteren van werk en het archiveren geen hoge prioriteit heeft voor ambtenaren omdat het veelal een extra handeling is, iets dat ook door geïnterviewden uit het archiefveld wordt onderschreven.¹¹⁶

Daarnaast is er vaak ook sprake van misverstanden over wat voor de lange termijn bewaard moet worden en welke vorm dat moet krijgen. Een geïnterviewde archivaris beschreef een veelvoorkomende situatie waarin ambtenaren ervan uitgaan dat stukken alleen in PDF/A -formaat overgedragen mogen worden, en dat alles dat niet in deze standaard om te zetten is dus niet 'archiefwaardig' is.¹¹⁷ Hierdoor ontstaat een situatie waarin juist de belangrijkste, meest complexe werkprocessen het slechtst gedocumenteerd worden, omdat de uitvoerders van deze processen aannemen dat het niet te doen is om deze documentatie effectief te beheren en de documentatie dus als vanzelf niet als 'archiefwaardig' wordt beschouwd.¹¹⁸ Een andere archivaris constateerde dat er vaak ook nog een verwachting is dat medewerkers van Documentaire Informatievoorziening (DIV)

¹¹¹ Geïnterviewde nr. 3 (gesprek op 3 april 2020); geïnterviewde nr. 5 (gesprek op 8 april 2020); geïnterviewde nr. 6 (gesprek op 8 april 2020); geïnterviewde nr. 8 (gesprek op 8 april 2020)

¹¹² Geïnterviewde nr. 6.

¹¹³ Geïnterviewde nr. 3.

¹¹⁴ Geïnterviewde nr. 1 (gesprek op 1 april 2020); geïnterviewde nr. 12 (gesprek op 24 april 2020); geïnterviewde nr. 13 (gesprek op 24 april 2020).

¹¹⁵ Geïnterviewde nr. 12.

¹¹⁶ Geïnterviewde nr. 3; geïnterviewde nr. 7 (gesprek op 8 april 2020); geïnterviewde nr. 8.

¹¹⁷ Geïnterviewde nr. 4 (gesprek op 3 april 2020).

¹¹⁸ Ibidem.

of Records Management (RM) archivering en dossiervorming voor hun rekening nemen, terwijl de systemen en processen erop zijn ingericht dat ambtenaren dit zelf doen.¹¹⁹

Deze laatste geïnterviewde zag de gevolgen hiervan terug bij inspecties, waarbij geregeld bleek dat processen als het aanvragen van vergunningen niet meer volledig te reconstrueren waren. In sommige gevallen was het zelfs niet mogelijk om de vergunning zelf terug te vinden.¹²⁰ Hieruit blijkt al wel hoe groot de gevolgen kunnen zijn als de documentatie en archivering niet op orde zijn: het wordt op deze manier onmogelijk om vast te stellen hoe de toekenning (of weigering) van de vergunning was verlopen, waardoor latere toetsing hiervan eveneens onmogelijk wordt. Een archivaris die veel werkt met kleinere gemeenten, benoemde dat de overgang van papier naar digitaal vaak een probleem is, omdat gemeenten zich er vaak nog onvoldoende bewust van zijn dat digitaal informatiebeheer om hele andere vaardigheden vraagt dan op papier.¹²¹ Aangezien het bestaande informatiebeheerpersoneel veelal onvoldoende getraind is in deze vaardigheden,¹²² kunnen er grote gaten ontstaan in de gevormde dossiers.

Enkelen van de gesproken archivariissen omschrijven wel hoe ze proberen dit te verbeteren door zich meer te profileren binnen hun organisatie. Uitgangspunt hierbij is vrijwel altijd de 'archiving by design'-aanpak: zorgen dat de voorwaarden die nodig zijn voor het goed archiveren en duurzaam toegankelijk houden van informatie, in een zo vroeg mogelijk stadium in een werksysteem aanwezig zijn.¹²³ Zaak hierbij is, zo blijkt uit de gesprekken, om ook zo vroeg mogelijk bij het aankoopproces betrokken te raken. Idealiter zijn de processen zo ingericht dat dit vanzelf gebeurt, maar vaak vraagt dit ook om een actieve houding van archiefprofessionals om te zorgen dat ze niet pas achteraf advies uitbrengen.¹²⁴ Een geïnterviewde gemeentearchivaris vertelde hoe het jaren kostte om digitaal informatiebeheer op de radar te krijgen, door dit onderwerp zoveel mogelijk aan te kaarten en door gebruik van de woorden 'archief' of 'archivaris' te vermijden, in dat geval 'sta je 2-0 achter', aldus deze geïnterviewde.¹²⁵

Informatiebeheer bij grote en kleine gemeenten

Dit bevestigt wederom het idee dat archivering geen hoge prioriteit heeft binnen gemeentelijke organisaties. Er zijn ook wel pogingen gedaan om het bewustzijn van het belang van goed

¹¹⁹ Geïnterviewde nr. 6.

¹²⁰ Geïnterviewde nr. 6.

¹²¹ Geïnterviewde nr. 8.

¹²² Geïnterviewde nr. 7.

¹²³ Nationaal Archief, *Archiveren by design*, website Nationaal Archief: <https://www.nationaalarchief.nl/archiveren/kennisbank/archiveren-by-design>. Laatst gewijzigd op: onbekend. Laatst geraadpleegd op 7 mei 2020.

¹²⁴ Geïnterviewde nr. 3; geïnterviewde nr. 5; geïnterviewde nr. 7; geïnterviewde nr. 8.

¹²⁵ Geïnterviewde nr. 11 (gesprek op 16 april 2020).

informatiebeheer en goede archivering te verhogen, bijvoorbeeld door eisen op dit gebied vast te leggen in de Gemeentelijke Inkoopvoorwaarden bij IT (GIBIT), maar deze eisen worden vaak genegeerd in de uitvoering.¹²⁶ Zoals wel is gebleken uit de analyse in hoofdstuk 2 van de richtlijnen van het Ministerie van Justitie en Veiligheid, kan gebrekkig digitaal informatiebeheer onmogelijk samengaan met eerlijk, uitlegbaar, transparant en verantwoord gebruik van algoritmen door de overheid. Immers, wanneer bijvoorbeeld een besluitproces waarbij een algoritme betrokken was, niet meer volledig te reconstrueren is (zoals bijvoorbeeld de vergunningen uit het eerder genoemde voorbeeld), kan onmogelijk aan deze eisen worden voldaan.

Aangezien ook de rol van algoritmen in het handelen van gemeenten gereconstrueerd moet kunnen worden, is het goed om nog dieper in te gaan op de problemen met archivering en informatiebeheer bij gemeenten. Als er een ding duidelijk wordt uit de interviews, is dat wel hoe sterk de kwaliteit hiervan uiteen kan lopen per gemeente. Tekenend hiervoor waren twee interviews, beiden afgenomen op 8 april 2020. De eerste geïnterviewde, werkzaam voor een grote gemeente, maakte deel uit van een team dat ingesteld is om bij elk nieuw project waarbij nieuwe systemen aangeschaft en ingericht worden, te zorgen dat het informatiebeheer bij deze systemen op orde is.¹²⁷ De tweede geïnterviewde werkte bij een archiefdienst waarbij diverse kleinere gemeenten zijn aangesloten, en vertelde hoe in veel van deze gemeentes nog altijd het idee overheerste dat archivering over papier ging, ook al werd er inmiddels grotendeels digitaal gewerkt.¹²⁸ Dit wordt onderschreven door de ervaringen van een andere geïnterviewde, die stelde dat kleinere gemeenten vaak niet de expertise in huis hebben om goed digitaal informatiebeheer in te richten.¹²⁹

Dit wil natuurlijk niet zeggen dat in de grotere gemeenten alles perfect geregeld is. Als de nieuwe Archiefwet wordt ingevoerd, en de overbrengingstermijn wordt ingekort van twintig naar tien jaar – waarmee dus ook plotseling veel meer digitaal materiaal overgebracht zal moeten worden – zullen hier ongetwijfeld de nodige problemen naar voren komen, aangezien ook zij de moeilijke transitie naar digitaal werken hebben moeten maken. Met de overbrenging is het werk echter nog niet klaar: vervolgens is het de taak van de archiefdienst waaraan de stukken zijn overgedragen, om deze te verwerken, op te slaan en beschikbaar te stellen, in het geval van sommige stukken voor altijd. Verschillende geïnterviewden spraken uit dat ze verwachtten dat deze preservering een grote uitdaging zou worden, onder meer vanwege de complexiteit van digitale dossiers (bijvoorbeeld door het gebruik van veel verschillende bestandsformaten) en het dynamische karakter van digitale informatieobjecten.¹³⁰

¹²⁶ Geïnterviewde nr. 3; geïnterviewde nr. 4.

¹²⁷ Geïnterviewde nr. 7.

¹²⁸ Geïnterviewde nr. 8.

¹²⁹ Geïnterviewde nr. 3.

¹³⁰ Geïnterviewde nr. 5; geïnterviewde nr. 8; geïnterviewde nr. 11.

Hiermee komen we eigenlijk ook weer terug bij een eerder punt, namelijk de positionering van archieftaken binnen een organisatie. Veel problemen komen pas naar voren bij de overdracht van stukken aan de archiefdienst, wanneer het eigenlijk al te laat is om hier iets aan te doen en waardoor de archiefdiensten geconfronteerd worden met conserveringsproblemen. Deze zouden er altijd al zijn, maar het wordt niet makkelijker als de stukken in onvolledige, slecht geordende of onvoldoende beschreven staat aangeleverd worden. Dit sluit aan bij opmerking van een geïnterviewde, die concludeerde dat het archiefwezen ‘de boot gemist’ heeft als het gaat om de aansluiting op de gemeentelijke informatiehuishouding, doordat deze twee velden in Nederlandse overheidsorganisaties altijd gescheiden zijn gebleven.¹³¹ De scheiding is alleen maar groter geworden door de overgang naar digitaal werken, waarbij archieven vooral door zijn gegaan met het verwerken van de papieren werkvoorraden.

Een probleem dat hier nog bijkomt, is dat gemeenten en andere overheidsorganisaties nieuwe samenwerkingsverbanden opzetten, veelal ten behoeve van efficiëntie, waarbij vaak geen duidelijke afspraken over informatiebeheer voor de lange termijn zijn gemaakt. Een voorbeeld hiervan, aangedragen door een van de geïnterviewden, zijn de Regionale Informatie en Expertise Centra (RIECs), samenwerkingen tussen overheidsorganen op verschillende niveaus, ter ondersteuning van de aanpak van georganiseerde misdaad.¹³² In een rapport over de RIECs constateren de Erfgoedinspectie, het Stadsarchief Rotterdam en het Interbestuurlijk Toezicht Noord-Brabant dat bij RIECs onduidelijk is wie verantwoordelijk is voor archivering in deze zogeheten ketensamenwerking, omdat mogelijk de verschillende partners ieder voor een eigen deel van de informatie verantwoordelijk zijn.¹³³ Dit is ook relevant voor het gebruik van algoritmen, aangezien kleinere gemeenten mogelijk in samenwerkingsverband software zullen laten ontwikkelen of analyses laten uitvoeren waar ze zelf niet de middelen voor hebben. Een geïnterviewde beschreef al een dergelijk systeem, waarin meerdere gemeentes samen bijdroegen aan een zelfstandig regionaal kenniscentrum voor data science en analyse.¹³⁴ Medewerkers van dit centrum werken volgens de geïnterviewde weliswaar altijd binnen teams van een gemeente, maar er blijft een risico bestaan dat er onduidelijke constructies gebruikt worden.

Algoritmeprojecten bij gemeenten

Het eerste wat naar voren komt bij het bespreken van lopende projecten met algoritmen, is dat men benadrukt dat het werk veelal nog in de testfase zit. Dit is echter lang niet in elke gemeente zo,

¹³¹ Geïnterviewde nr. 3.

¹³² Erfgoedinspectie, Stadsarchief Rotterdam en Interbestuurlijk Toezicht Noord-Brabant, *Integrale casusaanpak door de RIECs: Duurzame toegankelijkheid van archieven in de informatieketen* (Den Haag 2019) 5.

¹³³ Ibidem, 16-17.

¹³⁴ Geïnterviewde nr. 13.

aangezien uit een onderzoek uit 2018 bleek dat meerdere gemeenten al gebruik maakten van algoritmen.¹³⁵ De algoritmen zijn dus veelal nog niet ‘in productie’ genomen of onderdeel van officieel beleid, maar vooral testproducten. Als een algoritme wel volledig in gebruik is genomen, gaat het vrijwel altijd om een product van een derde partij, een situatie die hieronder verder besproken zal worden. Er kunnen diverse redenen zijn om de mogelijkheden van algoritmen te willen verkennen. Geïnterviewden noemden onder meer de wens om een reeds bestaand werkproces kwalitatief te verbeteren, evenals het effectiever of efficiënter maken van werkprocessen.¹³⁶ Zo noemde een geïnterviewde beleidsmedewerker de behoefte aan het verbeteren van detectie van bijstandsfraude als onderliggende reden voor het gebruiken van een algoritme: de bedoeling is dat het algoritme een extra ‘signaal’ afgeeft voor waar mogelijk sprake van fraude is, zodat de sociale recherche hier nader onderzoek naar kan doen.¹³⁷ De bedoeling hiermee is dus dat de beschikbare middelen efficiënter ingezet kunnen worden, een veelgebruikte toepassing van algoritmische data-analyse.¹³⁸

Uit de interviews blijkt dat het zogeheten sociale domein, waaronder ook de bijstand valt, veel genoemd wordt als mogelijke plek om algoritmen in te zetten.¹³⁹ Een mogelijke reden hiervoor werd gegeven door een geïnterviewde, werkzaam bij een afdeling Onderzoek, Informatie en Statistiek: fraudedetectie is een intensief proces, waar veel tijd en geld in gaat zitten, maar waarvan de uitvoerders tegelijkertijd brede bevoegdheden hebben om data te verzamelen.¹⁴⁰ Hierdoor is dus veel data beschikbaar en een grote behoefte om hier meer uit te halen, in de hoop dat het proces er efficiënter door wordt. Een andere geïnterviewde was sceptisch over dit idee van groeiende efficiëntie, aangezien veel gemeenten experimenteren met algoritmen, zonder dat de basale datahuishouding op orde is.¹⁴¹ Hierdoor is vaak nog niet goed in beeld welke gegevens er überhaupt beschikbaar zijn en wat de kwaliteit daarvan is, wat het lastig kan maken om betrouwbare analyses van deze gegevens te maken.

¹³⁵ S. Doove & D. Otten, *Verkenkend onderzoek naar het gebruik van algoritmen binnen overheidsorganisaties*. (Den Haag: CBS 2018). Te raadplegen op: <https://www.cbs.nl/nl-nl/maatwerk/2018/48/gebruik-van-algoritmen-door-overheidsorganisaties>. Laatst geraadpleegd op: 29 juli 2020; Anne Fleur van Veenstra e.a., *Quick scan AI in de publieke dienstverlening* (TNO, 2019). Te raadplegen op: <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/04/08/quick-scan-in-de-publieke-dienstverlening/TNO+-+AI+in+de+overheid+eindpresentatie+08042019.pdf>. Laatst geraadpleegd op 29 juli 2020.

¹³⁶ Geïnterviewde nr. 1; geïnterviewde nr. 2; geïnterviewde nr. 9; geïnterviewde nr. 10; geïnterviewde nr. 12.

¹³⁷ Geïnterviewde nr. 10, gesprek op 16 april 2020.

¹³⁸ Cathy O’Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York 2017).

¹³⁹ Geïnterviewde nr. 1, geïnterviewde nr. 2.

¹⁴⁰ Geïnterviewde nr. 2.

¹⁴¹ Geïnterviewde nr. 13

Zoals gezegd, hebben weinig gemeenten zelf de middelen in huis om complexere algoritmen te ontwikkelen. Een geïnterviewde werkt daarom samen met een plaatselijke universiteit aan de ontwikkeling van een algoritmetoepassing voor fraudedetectie in het sociaal domein, gecombineerd met een promotieonderzoek.¹⁴² Volgens de geïnterviewde verliep dit project in nauwe samenspraak met de sociale recherche, ook om van hun praktijkkennis gebruik te kunnen maken, en werd alleen gebruik gemaakt van gegevens die al in dit kader verzameld mochten worden, en van historische data van bevestigde fraudeurs. Hierover zei de geïnterviewde dat rechercheurs zich doorgaans baseerden op hun eigen ervaringen uit het verleden om te bepalen wie verder onderzocht werd.¹⁴³ Het is dus niet onmogelijk dat dergelijke historische biases toch in het algoritme terechtkwamen, al is het ook mogelijk dat het algoritme nieuwe patronen in de bestaande data zou vinden.

Er was bij dit project wel zicht op mogelijke ongewenste discriminatie, en ook uitlegbaarheid kreeg een belangrijke plaats. Zo werd ten bate van de uitlegbaarheid bij de eerste praktijktest een eenvoudiger ‘beslisboom’-algoritme gebruikt, terwijl bij het wetenschappelijk onderzoek wel gebruik gemaakt werd van complexere, algoritmen. Opvallend was dat de eisen op het gebied van documentatie bij het wetenschappelijk onderzoek hoger lagen dan bij de gemeente, en dat de universiteit op veel meer gebieden om expliciete uitleg en verantwoording vroeg.¹⁴⁴ Dit was vooral het gevolg van de gehanteerde wetenschappelijke benadering, die om gedetailleerde verantwoording vroeg, terwijl de gemeente meer gericht was op het unnen uitleggen van het concrete praktijkgebruik van het algoritme. Het is dus mogelijk dat de samenwerking met de universiteit in dit geval heeft gezorgd voor betere documentatie en bewustwording van welke keuzes en afwegingen uitgelegd moeten worden binnen een dergelijk project.

Uitlegbaarheid

In de interviews was ook uitlegbaarheid een belangrijk thema. Deze term wordt gebruikt in de richtlijnen van het Ministerie van Justitie en Veiligheid, maar hierin zit de nodige ruimte voor interpretatie (zie hoofdstuk 3). Vanwege deze openheid van het begrip en de relatief recente introductie ervan, werd uitlegbaarheid in de interviews nadrukkelijker besproken dan de andere criteria, om van alle geïnterviewden een idee te krijgen hoe zij deze term interpreteerden en hoe zij dachten over hoe in de praktijk aan deze eis voldaan zou kunnen worden. Het was ook het begrip dat verreweg de meeste reacties losmaakte bij de geïnterviewden. De uitlegbaarheidseis is immers een belangrijk onderdeel van het raamwerk voor verantwoord algoritmegebruik door overheden, aangezien deze voor een belangrijk deel omschrijft hoe algoritmegebruik naar burgers toe

¹⁴² Geïnterviewde nr. 2.

¹⁴³ Geïnterviewde nr. 2.

¹⁴⁴ Geïnterviewde nr. 2.

verantwoord moet worden. Daarnaast is het aannemelijk dat ideeën over uitlegbaarheid zullen raken aan archivering: de stukken waarin het algoritmegebruik wordt uitgelegd, zullen ook op de lange duur goed beschikbaar moeten zijn voor burgers. Verder moet de werking van het algoritme goed uitgelegd kunnen worden, waarbij onvermijdelijk het documenteren van het werk rondom de bouw van het algoritme een rol gaat spelen.

In dit onderwerp kwamen dus beide categorieën geïnterviewden samen. Een van de eerste kwesties die hierbij naar voren kwam, was de vraag of er ten behoeve van de uitlegbaarheid aparte stukken zouden moeten worden opgesteld. Een geïnterviewde, werkzaam in het informatiebeheer, stelde dat het makkelijker zou zijn als uitlegbaarheid niet om het opstellen van nieuwe stukken zou vragen, maar dat dit waarschijnlijk onvermijdelijk zou zijn.¹⁴⁵ Veel van het bestaande materiaal, in het bijzonder de broncode, is immers niet voldoende begrijpelijk voor de gemiddelde burger. Enkele geïnterviewden die aan projecten met algoritmen werkten, zagen weinig bezwaar tegen het opstellen van nieuwe stukken ten behoeve van de uitlegbaarheid: zij zagen dit als een belangrijk onderdeel van het gebruik van algoritmen door de overheid.¹⁴⁶ Daar staat echter tegenover dat een andere geïnterviewde de uitlegbaarheidseis als ‘oneerlijk’ omschreef, aangezien het in detail uitleggen van de werking van een algoritme een bijzonder lastige opgave is.¹⁴⁷

Dit legt direct de centrale vraag bloot die de uitlegbaarheidseis oproept: hoe ver moet de uitleg gaan, zeker in het geval van complexere algoritmen? Een geïnterviewde sprak zich hier heel duidelijk over uit: een ‘black box’ algoritme – oftewel een algoritme waarvan de interne werking niet duidelijk is – mag nooit gebruikt worden in een situatie waarin het burgers raakt.¹⁴⁸ Een archiefprofessional analyseerde het op een manier die hier goed op aansluit: naast uitlegbaar zou het algoritme toetsbaar moeten zijn, zodanig dat het in ieder geval door specialisten volledig begrepen en getoetst zou kunnen worden.¹⁴⁹ Deze geïnterviewde trok een vergelijking met wetteksten, die veelal voor leken ook niet goed te begrijpen zijn, maar die door specialisten geïnterpreteerd en uitgelegd kunnen worden. Het algoritme zou dan in minder detail uitgelegd hoeven worden, maar zou wel volledig open moeten zijn, zodat het voldoende getoetst kan worden.¹⁵⁰ Dit laat zien dat aan het idee van uitlegbaarheid veel facetten zitten, en dat het op verschillende manieren kan worden ingevuld. Een constante blijft echter de gedachte dat met het gebruik van algoritmen als overheid een zekere verplichting gepaard gaat om dit naar burgers toe te kunnen verantwoorden en verklaren.

¹⁴⁵ Geïnterviewde nr. 7.

¹⁴⁶ Geïnterviewde nr. 2; Geïnterviewde nr. 9.

¹⁴⁷ Geïnterviewde nr. 10

¹⁴⁸ Geïnterviewde nr. 9.

¹⁴⁹ Geïnterviewde nr. 11.

¹⁵⁰ Ibidem.

Daarnaast blijft nog de vraag staan hoe goed het werk dat waarschijnlijk ten behoeve van de uitlegbaarheid gedaan zal moeten worden, in bestaande werkvormen past. Een geïnterviewde vertelde dat er wel altijd korte verslagen en notities werden opgesteld van voltooide onderzoeken, maar dat deze niet gericht waren op het informeren van de burger, maar vooral op het inzichtelijk maken van het onderzoek voor collega's die het onderzoek nodig zouden kunnen hebben, of voor eigen hergebruik van het onderzoek.¹⁵¹ Deels wordt het werk dus al gedaan, maar niet gericht op de burger. Dit sluit aan bij wat een geïnterviewde archivaris beschreef als een cultureel probleem op het gebied van openbaarheidsdenken binnen Nederlandse gemeenten. Met het oog op de komende Wet Open Overheid (WOO), waarin overheidsorganisaties actief stukken openbaar moeten maken, besprak deze geïnterviewde de huidige Wet Openbaarheid van Bestuur (WOB): binnen deze wet is actieve openbaarmaking ook mogelijk, maar het wordt niet tot nauwelijks gedaan.¹⁵² Dit laat zien dat gemeenten meer zouden kunnen doen aan openbaarheid, waardoor het voldoen aan de uitlegbaarheidseis lastig zou kunnen worden.

Commercieel ontwikkelde algoritmen in gemeenteland

Een kwestie die nauw verbonden is met uitlegbaarheid, is de samenwerking met commerciële partijen. Zoals het geval van COMPAS (zie hoofdstuk 1) laat zien, is het namelijk lastig om de werking van algoritmen inzichtelijk te maken, als de maker van het algoritme zich op bedrijfsgeheimen en intellectueel eigendomsrecht kan beroepen. Het is echter wel een aantrekkelijke optie voor met name kleinere gemeenten, aangezien deze zelf vaak niet de middelen hebben om (complexe) algoritmen te ontwikkelen.¹⁵³ Hiermee ontstaat dus een duidelijk risico, namelijk dat gemeenten die gebruik willen maken van de mogelijkheden die algoritmische data-analyse biedt, hiervoor massaal de diensten van commerciële bedrijven inkopen, zonder dat er duidelijke afspraken zijn over inzicht in het gebruikte algoritme. Daarmee zou een situatie ontstaan waarin gemeenten beslissingen nemen met ondersteuning van algoritmen, zonder dat duidelijk is hoe de output van het algoritme tot stand is gekomen. Dit ondermijnt de legitimiteit van de beslissing en maakt het voor burgers onmogelijk om goed te begrijpen hoe, bijvoorbeeld, de beoordeling van hun vergunnings- of uitkeringsaanvraag tot stand is gekomen.¹⁵⁴

Enkele geïnterviewden gaven aan betrokken te zijn (geweest) bij projecten waarbij gebruik gemaakt werd van een algoritme dat ontwikkeld is door een commerciële partij.¹⁵⁵ Bij een geïnterviewde was sprake van een samenwerking die inmiddels al enkele jaren duurde en waarbij

¹⁵¹ Geïnterviewde nr. 12.

¹⁵² Geïnterviewde nr. 8.

¹⁵³ Geïnterviewde nr. 1; geïnterviewde nr. 6; geïnterviewde nr. 10.

¹⁵⁴ Danaher, 'The threat of algocracy', 249-255.

¹⁵⁵ Geïnterviewde nr. 1; geïnterviewde nr. 10; geïnterviewde nr. 12.

zich een vast proces gevormd heeft, dus deze casus zal hier als voorbeeld dienen. Het gaat in dit geval om een toepassing in de detectie van bijstandsfraude. De gemeente levert eens per kwartaal een dataset aan het data analytics bedrijf dat het algoritme ontwikkeld heeft. De dataset is een gepseudonimiseerde afgeleide van de gegevens die de gemeente van haar bijstandsgerechtigde populatie heeft, waaruit naam, adres en woonplaats verwijderd zijn. Het algoritme, dat getraind is met historische data van fraudeurs, wijst op basis van de dataset aan ieder individu een risicoscore toe, waarbij afwijkende datapunten meegerekend worden, evenals waarden die historisch gezien in verband zouden staan met fraude. Bij de risicoscores staat verder altijd een duiding van hoe de score tot stand is gekomen, evenals drie 'subscores' die het eindtotaal onderbouwen. Handhavers bij de gemeente de-pseudonimiseren de lijst en zoeken naar de tien hoogste scores die nog niet eerder onderzocht zijn. De hoge risicoscore heeft de status van een 'signaal', vergelijkbaar met een tip van een buurtbewoner, wat aanleiding is tot een vooronderzoek naar fraude, en eventueel een vervolgonderzoek als het vooronderzoek meer aanwijzingen van fraude oplevert.¹⁵⁶

Duidelijk is echter dat ook in deze gemeente nog dilemma's zijn, bijvoorbeeld als het aankomt op het bewaren van de uitkomsten: tot voor kort werd het bestand met risicoscores om privacyredenen weggegooid, maar inmiddels is besloten het voorlopig te bewaren, in het kader van controleerbaarheid, in ieder geval totdat de bezwaartermijn verstreken is.¹⁵⁷ Dit maakt het in ieder geval mogelijk om de output van het algoritme te controleren, al hebben bijstandsgerechtigden tot nu toe nog geen vragen gesteld over het gebruik van het algoritme. Het algoritme zelf blijft echter eigendom van het bedrijf, dat technische documentatie over de werking van het algoritme aan de gemeente ter beschikking stelt.

Een andere geïnterviewde, werkzaam als zelfstandige en als adviseur bij gemeenten op het gebied van data en AI, ging uitgebreid in op de risico's die samenwerking met commerciële partijen met zich mee kan brengen, met name als een gemeente zelf nog onvoldoende bekend is met datagedreven werken.¹⁵⁸ Volgens deze geïnterviewde ontbreekt het in veel gemeenten namelijk nog aan kennis en ervaring op dit gebied, waardoor ambtenaren zich nog onvoldoende bewust zijn van de risico's die dergelijke samenwerkingen met zich mee brengen. Zo kwam er een voorbeeld ter sprake van een project waarbij een gemeente persoonlijke data door een externe partij liet analyseren. Bij nader onderzoek hiernaar, in het kader van een WOB-verzoek, kwam naar voren dat er onvoldoende vragen waren gesteld over informatiebeveiliging en de inhoudelijke details van de verwerking. Een verwerkingsovereenkomst ontbrak zelfs volledig. Dit sluit aan bij bevindingen van een andere geïnterviewde, die stelde dat het nodig was om de kennis over het werken met data, en

¹⁵⁶ Geïnterviewde nr. 10.

¹⁵⁷ Ibidem.

¹⁵⁸ Geïnterviewde nr. 13.

de risico's die hieraan verbonden zijn, eerst te verbeteren in de organisatie. Zodra deze kennis, wel aangeduid als 'datageletterdheid' of 'datavakmanschap', op peil is, kan er verantwoord mee gewerkt worden.¹⁵⁹

Het is inmiddels tijd om te kijken naar wat er over al deze thema's gezamenlijk te zeggen is. Uit de gesprekken over archivering en informatiebeheer is duidelijk geworden dat er nog het nodige verbeterd kan worden op deze gebieden bij gemeenten. Het algemene beeld dat op dit gebied ontstaat, is dat archivering lage prioriteit heeft bij gemeenten, en dat archiefdiensten moeite hebben om zich zichtbaar te maken in de organisatie. Hierop zijn echter ook uitzonderingen en er liggen zeker ook kansen: nu de overgang naar digitaal werken in het algemeen wel volbracht is, groeit bij sommige gemeenten al meer het besef van het belang van goed informatiebeheer en goede archivering. Dit vraagt echter wel om bewuste inspanningen van archief- en informatieprofessionals, om hun werk zichtbaar te maken en het belang ervan te tonen. Op basis van de interviews wordt duidelijk dat er grote risico's ontstaan als hier onvoldoende oog voor is, zoals het ontstaan van incomplete dossiers, waardoor processen niet meer te reconstrueren zijn, zeker wanneer er gewerkt wordt binnen nieuwe samenwerkingsorganisaties. Dit brengt grote schade toe aan de verantwoording en transparantie van het overheidshandelen, aangezien dit met incomplete archieven nooit goed gereconstrueerd kan worden.

Bij de algoritmeprojecten benadrukken de geïnterviewden vooral dat deze doorgaans nog in een vroege, experimentele fase zitten. Hieruit blijkt nog meer hoe belangrijk het is om nu, in dit vroege stadium, te zorgen dat eerlijkheid, transparantie en verantwoording geborgd blijven. Verder laat het ook zien dat in veel gevallen de kennis nog opgebouwd moet worden binnen gemeenten, die daarom vaak samenwerkingen opzoeken met andere gemeenten, universiteiten of commerciële bedrijven. Dat het als experimenteel wordt gezien, wil echter natuurlijk niet zeggen dat het geen invloed zal hebben op beleid. Er lijkt een opvatting te heersen dat alleen zaken die op het hoogste niveau besproken zullen worden, op een 'officiële' manier moeten worden gedocumenteerd en vastgelegd. Hierdoor ontstaat het risico dat veel informatie over de beginfase van het ontstaan van werkprocessen met algoritmen niet goed wordt vastgelegd en daardoor niet goed bewaard zal worden. Daarmee kan dit proces niet meer gereconstrueerd worden en wordt het erop gebaseerde beleid ondoorzichtiger. Samenwerkingen met commerciële partijen zijn een begrijpelijke stap, maar brengen het gevaar met zich mee dat cruciale informatie over de werking van algoritmen wegens bedrijfsbelangen verborgen blijft. Het is dan ook belangrijk dat hier goede afspraken over gemaakt worden en dat ambtenaren zich bewust worden van de risico's die dit werk met zich meebrengt.

¹⁵⁹ Geïnterviewde nr. 9.

Zoals gezegd betekent dit ook dat het informatiebeleid op orde moet zijn en dat er gezorgd moet worden voor volledige dossiervorming, anders verdwijnt een van de pijlers van het democratische systeem: een controleerbare overheid.

Hoofdstuk 4: Synthese en aanbevelingen

Op basis van de informatie uit de vorige hoofdstukken, kunnen nu aanbevelingen geformuleerd worden. Het uitgangspunt voor deze aanbevelingen, is het beantwoorden van de vraag ‘Welke veranderingen zijn in de praktijk nodig om eerlijk, verantwoordelijk, transparant en uitlegbaar gebruik van algoritmen door de overheid mogelijk te maken met archivering en informatiebeleid?’. In zekere zin loopt dit natuurlijk een beetje vooruit op de conclusie, maar op deze manier kunnen de aanbevelingen in voldoende detail besproken worden.

Ten eerste zullen drie randvoorwaarden worden besproken die meer op strategisch niveau zitten: interdisciplinaire samenwerking, verbeteringen van de standaarden en archiveringssystemen, en de informatiecultuur. Deze drie vlakken vormen de voornaamste randvoorwaarden die het archiverings- en informatiebeleid in een organisatie geschikt moeten maken voor het bijdragen aan een wenselijke vorm van algoritmegebruik door overheden. Vervolgens komen wat meer praktische aanbevelingen aan bod, die aan de hand van twee doelstellingen behandeld worden: het vastleggen van de context van de ontwikkeling en het gebruik van algoritmen, en het borgen van uitlegbaarheid. Dit waren beiden belangrijke thema's in de literatuurstudie en de interviews. Het zijn beiden natuurlijk nog abstracte principes, maar de aanbevelingen worden verder geconcretiseerd en uitgewerkt om invulling aan deze principes te geven.

Interdisciplinaire problematiek

In de eerste plaats is het belangrijk om te kijken naar de randvoorwaarden die nodig zijn om op een eerlijke, transparante, verantwoordelijke en uitlegbare manier met algoritmen te werken. Een kwestie die tijdens de interviews duidelijk naar voren kwam, was dat het ontbrak aan samenwerking en overleg tussen disciplines. Dit was het duidelijkst als het ging om de betrokkenheid van archief- en informatieprofessionals bij algoritmeprojecten: dit kwam vrijwel nooit voor. De archief- en informatieprofessionals stelden dat ze in het algemeen vaak te laat of helemaal niet betrokken werden bij projecten of bij zaken als de inkoop van nieuwe applicaties. Verder constateerden ze ook dat er regelmatig misverstanden waren over wat er gearchiveerd moest worden en wie er verantwoordelijk was voor het in goede staat houden van het digitale archief. Aan de andere kant vonden de medewerkers van algoritmeprojecten dat archiefdiensten vaak achter de feiten aan liepen of weinig betrokken waren.

Hiermee blijft niet alleen essentiële expertise buiten het proces, er ontstaat ook een groot risico dat het werk van de overheid onvoldoende gereconstrueerd kan worden doordat de bewaarde documentatie onvoldoende is. Hierdoor kunnen wetten als de Archiefwet en de WOO niet worden nageleefd, kan een overheidsorganisatie onvoldoende verantwoordelijk worden gehouden en wordt

er ook te weinig gedaan om transparantie en uitlegbaarheid veilig te stellen. Een multidisciplinaire overlegstructuur voor het gebruik van algoritmen zou hier, maar ook op andere vlakken, erg helpend kunnen zijn. Zoals ook blijkt uit het Europese kader voor vertrouwenswaardige, *human-centric* AI, zijn er tal van facetten die komen kijken bij het vaststellen van wat een wenselijke implementatie van AI zou zijn. AI kan zo breed ingezet worden en zulke brede gevolgen hebben, dat er rekening gehouden moet worden met ethische, juridische, technologische, politieke, maatschappelijke en economische vraagstukken. Het is de brede, multidisciplinaire aard van het werken met algoritmen, die om even brede en multidisciplinaire samenwerkingsverbanden vraagt.

Op landelijk niveau is er al in zekere zin een dergelijk overlegstructuur, de Nederlandse AI Coalitie.¹⁶⁰ Binnen de AI coalitie overleggen instellingen uit overheid, wetenschap en bedrijfsleven over de toepassing van AI in Nederland. Echter zou het ook goed zijn om op meer lokaal niveau dergelijke overleggen in te stellen, aangezien hier ook beslissingen worden genomen en experimenten worden opgestart waarbij AI een belangrijke rol speelt. Hierin zouden experts uit verschillende disciplines zitting moeten hebben en werken aan het ontwikkelen van duidelijke kaders en standaarden. Deze kunnen dan al voorafgaand aan de start van een project toegepast worden, in plaats van dat er achteraf ingegrepen moet worden als er iets fout gaat. Deze groep kan ook zorgen dat richtlijnen die op landelijk of zelfs Europees niveau opgesteld zijn, vertaald kunnen worden naar de praktijk van, bijvoorbeeld, het gemeentelijk niveau.

In een dergelijke groep zouden in de eerste plaats natuurlijk mensen met technische expertise vertegenwoordigd moeten zijn: de programmeurs die algoritmen kunnen maken en de technische aspecten die hierbij komen kijken kunnen begrijpen. Daarnaast zullen ook data-analisten vertegenwoordigd moeten zijn, bijvoorbeeld vanuit een afdeling Onderzoek & Statistiek of analytics teams. Een groep die zeker in dit onderzoek niet weggelaten mag worden, zijn natuurlijk de archief- en informatieprofessionals. Zij kunnen meedenken over effectief databeheer en hebben kennis over hoe documentatie effectief en duurzaam bewaard kan worden – essentieel voor het mogelijk maken van transparantie, verantwoording en uitlegbaarheid. Verder is het natuurlijk ook belangrijk dat de privacyfunctionaris aanwezig is bij deze groep, gezien de hoeveelheid en gevoeligheid van gegevens waarmee gewerkt wordt. Beleidsmedewerkers en uitvoerders uit het domein waarin een specifiek algoritme gebruikt wordt, mogen ook niet ontbreken. Een burgervertegenwoordiging zou ook passend zijn, mogelijk in samenspraak met lokale organisaties of grotere NGO's. Tot slot zouden ook juridische medewerkers hierbij betrokken moeten zijn, bijvoorbeeld wanneer het gaat om intellectueel eigendomsrecht.

¹⁶⁰ Meer informatie over de AI Coalitie is te vinden op de website: <https://nlaic.com/>.

Hogere standaarden

Een andere belangrijke voorwaarde, is het aanscherpen van de standaarden voor metadatering en opslag van documenten. Uit de interviews is duidelijk geworden dat er voor stukken die op politiek niveau behandeld worden, bijvoorbeeld in het college of de raad, doorgaans wel een infrastructuur bestaat om deze stukken goed en geordend te bewaren, veelal in een Document Management Systeem (DMS) of zaakstelsel. De interviews lieten echter ook zien dat bestanden die buiten dit niveau blijven, vaak met weinig extra informatie bewaard worden in persoonlijke of gezamenlijke netwerkmappen. Hierbij wordt veelal gebruik gemaakt van applicaties als SharePoint, dat als onderdeel van het populaire pakket Office 365 onder meer cloudopslag en documentbeheer aanbiedt.

Dit kan snel problemen opleveren, vooral in combinatie met de houding, die ook in de interviews naar voren kwam, dat alleen stukken die naar het DMS gaan als ‘officieel’ gelden. De rest is vooral voor eigen gebruik en heeft verder, in de ogen van medewerkers, geen officiële status. Het kan hier echter wel gaan om stukken die relevant zijn voor bewijsvoering en voor reconstructie van de werkzaamheden, zoals onderzoeksverslagen. Bovendien is ‘officiële’ status, of invoer in een DMS, kennelijk geen garantie dat stukken goed bewaard blijven, zoals wel bleek uit de ervaring van een geïnterviewde archiefinspecteur, die constateerde dat vergunningen soms niet terug te vinden waren (zie hoofdstuk 3). Dit is met name problematisch bij algoritmen, omdat de context van de ontwikkeling ervan zo belangrijk kan zijn voor het begrijpen van het functioneren van het algoritme. Wanneer dus zoveel informatie verloren gaat, kan er een groot gat ontstaan in de kennis over de werking van het algoritme.

Dit laat verder een belangrijk probleem van digitaal werken zien: zodra stukken niet van het begin af aan op de juiste plek en met voldoende metagegevens zijn opgeslagen, kan het al snel onmogelijk worden om ze terug te vinden. Het verdient nog wel even aandacht om te kijken wat in dit geval ‘het begin’ is. Voorstanders van de ‘archiving-by-design’ aanpak stellen dat het in een digitale omgeving niet voldoende is om een stuk in een informatiebeheersysteem te ‘vangen’ (*to capture*, in de Engelstalige literatuur) zodra het aangemaakt wordt. Voor een echt gegarandeerd complete dossiervorming is het nodig om hier al maatregelen voor te treffen voordat het stuk wordt aangemaakt. Archiving-by-design houdt in dat bij de aankoop, bouw of aanpassing van een werksysteem wordt bepaald wat er nodig is om de informatie die in dat werksysteem terechtkomt, duurzaam toegankelijk te maken.¹⁶¹ Zo hoeven er achteraf geen extra maatregelen getroffen te worden in het kader van archivering.

¹⁶¹ Website Nationaal Archief, *Archiveren by design*. Te raadplegen op: <https://www.nationaalarchief.nl/archiveren/kennisbank/archiveren-by-design>. Laatst geraadpleegd op 26 augustus 2020; Website Nationaal Archief, *Archivering by design, lessons learned. Concept voor het ontwerpen*

De ‘capture’ van stukken is echter zinloos zonder dat er voldoende metadata meekomt. Metadata zorgt ervoor dat stukken te traceren en te contextualiseren zijn, zelfs als de inhoud van het stuk verloren is gegaan.¹⁶² Deels gaat het hier om ouderwetste metadata, te vergelijken met vormen van metadatering in het papieren domein, zoals het samenstellen van dossiers en mappen waarin documenten ondergebracht worden. Het begrip is echter aanzienlijk breder te trekken in het digitale domein, waar elke handeling sporen achterlaat in de vorm van veelal automatisch gegenereerde metadata. Het afvangen en vormgeven hiervan kan een enorme bijdrage leveren aan de traceerbaarheid van stukken en het reproduceren van de processen waarmee ze tot stand zijn gekomen.

Het is wel belangrijk om hierbij niet te hard van stapel te lopen, aangezien veel van de ‘papieren’ metadatavormen vaak ook nog onvoldoende worden bijgehouden. Er zijn hiervoor in de praktijk nauwelijks (of geen) formele richtlijnen, waardoor bestanden worden opgeslagen in mappen naar inzicht van de medewerker, of hoogstens de projectleider, zonder overkoepelend toezicht zoals afdelingen Documentaire Informatievoorziening (DIV) dat in het papieren tijdwerk deden. Dit zou dan ook de eerste stap moeten zijn, om de basis van de informatiehuishouding op orde te brengen. Er moeten duidelijke standaarden komen, liefst overkoepelend binnen Nederland of internationaal, voor hoe bestanden door werknemers worden opgeslagen en beschreven. Archiefdiensten zouden een belangrijke rol moeten spelen in het ontwikkelen van deze standaarden, aangezien zij de kennis over duurzame toegankelijkheid voor de lange termijn in huis hebben, of anders zouden moeten zorgen dat ze deze kennis krijgen.

Veranderingen in de informatiecultuur

Hierbij hoort echter ook dat er een cultuurverandering moet plaatsvinden, zodat het belang van een goede informatiehuishouding tot medewerkers doordringt. Anders kunnen er nog zoveel regels en standaarden ingevoerd worden, maar zullen ze niet of nauwelijks nageleefd worden.¹⁶³ In de archiefwetenschap wordt ‘informatiecultuur’ doorgaans gedefinieerd als het geheel aan waarden en praktijken rondom het gebruik van en de omgang met informatie.¹⁶⁴ Uit de interviews blijkt wel dat

van een informatiesysteem (26 oktober 2018). Te raadplegen op:

<https://www.nationaalarchief.nl/archiveren/nieuws/archivering-by-design-lessons-learned>. Laatst geraadpleegd op 26 augustus 2020.

¹⁶² Frank Upward, Barbara Reed, Gillian Oliver en Joanne Evans, *Recordkeeping Informatics for a Networked Age* (Clayton 2017) Hoofdstuk 8, te raadplegen op:

http://books.publishing.monash.edu/apps/bookworm/view/Recordkeeping+Informatics+for+a+Networked+Age/207/19_chapter-title-8.html.

¹⁶³ Ibidem, hoofdstuk 3. Te raadplegen op:

http://books.publishing.monash.edu/apps/bookworm/view/Recordkeeping+Informatics+for+a+Networked+Age/207/13_chapter-title-3.html.

¹⁶⁴ Ibidem.

er sprake is van een duidelijk informatiecultuurprobleem bij Nederlandse gemeenten. Zoals hierboven al benoemd, is er weinig sprake van overleg met archief- en informatiespecialisten en heerst er bij veel ambtenaren het idee dat zij niet verantwoordelijk zijn voor hoe door hen opgestelde documenten, worden bewaard. Er is dus sprake van een te 'losse' informatiecultuur, waarin onvoldoende waarde wordt gehecht aan volledig en correct informatiebeheer. Dit is natuurlijk een hele brede uitspraak, op basis van een beperkt aantal interviews. In werkelijkheid zal de informatiecultuur er voor iedere gemeente anders uitzien, maar in algemene zin lijkt dit het voornaamste probleem te zijn.

Gillian Oliver en Fiorella Foscari, twee experts op het gebied van informatiecultuur, schrijven over de diverse facetten van informatiecultuur en over mogelijke manieren om binnen een organisatie de informatiecultuur te beïnvloeden of te wijzigen.¹⁶⁵ Centraal in die benadering staat dat informatiecultuur nauw samenhangt met organisatiecultuur.¹⁶⁶ Deze zal dus in geen enkele organisatie precies hetzelfde zijn, waardoor het ook niet realistisch is om één generieke benadering voor alle Nederlandse gemeenten voor te stellen. Het doel is echter wel overal gelijk: de informatiecultuur wijzigen, of ten minste bijsturen, zodat er meer belang wordt gehecht aan het volledig en geordend bewaren van informatie. Zo blijft deze op de lange termijn vindbaar en toegankelijk, wat bijdraagt aan de borging van verantwoording en transparantie in een overheidsorganisatie, waardoor ook eerlijkheid beter gecontroleerd kan worden. Dit onderzoek bood geen ruimte om uitgebreid informatieculturen bij afzonderlijke gemeenten in kaart te brengen, aangezien reikwijdte van dit onderzoek een stuk breder was. Het zou wel waardevol vervolgonderzoek kunnen zijn om preciezer en diepgaander informatieculturen binnen Nederlandse gemeenten in kaart te brengen, zeker in het kader van algoritmegebruik, zodat hier nog specifiekere adviezen over geformuleerd kunnen worden.

Context veiligstellen met archivering-by-design

Het eerste uitgangspunt voor de meer praktische aanbevelingen, is dat de context waarin algoritmen worden ontwikkeld en gebruikt volledig en duurzaam wordt vastgelegd. Uit de in hoofdstuk 1 besproken literatuur wordt dit ook duidelijk. Auteurs als Introna, Reigeluth en Danaher hameren op het belang van context om verschillende redenen. Ten eerste gaat het om het proces waarin het algoritme gemaakt wordt: hierbij zijn vaak veel partijen betrokken, van een opdrachtgever tot vaak meerdere programmeurs en projectmanagers, die allemaal een gedeelde vorm van 'agency' hebben, waarin ze allemaal beslissingen en aannames kunnen maken die invloed hebben op de uiteindelijke

¹⁶⁵ Gillian Oliver en Fiorella Foscari, *Records Management and Information Culture : Tackling the People Problem* (London 2014).

¹⁶⁶ Ibidem, 35-37.

werking van het algoritme.¹⁶⁷ Ditzelfde geldt voor de datasets die als input- of trainingsdata gebruikt worden. De winning en samenstelling hiervan gaat ook gepaard met tal van bewuste en onbewuste keuzes en aannames, die zonder goede archivering onzichtbaar blijven.¹⁶⁸ Op abstracter niveau past het ook in de techniekfilosofische benadering van Verbeek: Verbeek pleit voor een kritische reflectie op het gebruik van technologieën, waarin mensen zelf kunnen vormgeven hoe technologie hun leven beïnvloedt.¹⁶⁹ Zonder een duidelijk beeld van waar en door wie de werking van technologieën (zoals algoritmen) beïnvloedt kan worden, is het voeren van een dergelijke discussie niet goed mogelijk.

Daarnaast is dit relevant vanuit het perspectief van de in hoofdstuk 2 besproken regelgeving en beleidskaders. Zowel bij de huidige als bij de komende vernieuwing van de Archiefwet is veel aandacht voor de rol die archieven hebben als ‘verantwoordingsinstrument’ van de overheid: het bewaren van archieven zorgt ervoor dat het overheidsbeleid controleerbaar is door middel van (historisch) onderzoek in de archieven. Hiermee wordt verzekerd dat overheidsorganisaties verantwoording afleggen over hun handelen, een onmisbaar element in een democratie. Als algoritmen onderdeel worden van het overheidshandelen, wat ze voor een deel misschien al zijn, moet dus ook hierover verantwoording kunnen worden afgelegd. Het algoritme zelf kan wellicht een set ‘simpele’ instructies volgen waar op zichzelf weinig aan te analyseren valt, maar ook het opstellen van de instructies is onderdeel van het overheidshandelen en moet daarom traceerbaar zijn.

Het ‘volledig en duurzaam vastleggen van de context waarin algoritmen worden ontwikkeld en gebruikt’, zoals de aanbeveling aan het begin van deze paragraaf is geformuleerd, is nog steeds een vrij abstracte eis. Om dit in de praktijk te implementeren, is in de eerste plaats het eerder genoemde archiving-by-design nodig. Dit zou concreet betekenen dat al in een vroege fase van het algoritmeproject advies ingewonnen zou moeten worden van de archiefdienst waarbij de gemeente is aangesloten, of een interne specialist met voldoende kennis van archiefwetgeving. Dit advies zou moeten gaan over de geschiktheid van de gewenste applicaties voor langdurige archivering en hoe deze geïntegreerd zijn, of kunnen worden, in het archiveringssysteem van de gemeente. Let wel dat hierbij natuurlijk ook een verantwoording ligt voor de archiefdienst, of de gemeente, om dergelijke expertise in huis te hebben. Naleving hiervan zou gecontroleerd kunnen worden door, bijvoorbeeld, de provincies vanuit hun bestaande rol van toezichthouder op gemeenten.

De bedoeling hiervan is dat niet alleen raadsstukken in een goed beheerd archiveringssysteem terechtkomen, maar dat ook de andere stukken voor de lange termijn bewaard kunnen blijven, zodat ze vindbaar kunnen blijven voor latere onderzoekers. Het is belangrijk dat hierover aan het begin van het project afspraken worden gemaakt, maar het is nog belangrijker dat

¹⁶⁷ Introna, ‘Algorithms’.

¹⁶⁸ boyd en Crawford, ‘Critical Questions’; Reigeluth, ‘Why data is not enough’.

¹⁶⁹ Verbeek, ‘The Struggle for Technology’; Verbeek, *Op de vleugels van Icarus*.

deze principes niet uit het oog verloren worden in de loop van het project. Aangezien continu toezicht op hoe documentatie wordt opgesteld en bewaard realistisch noch wenselijk is, zal dit waarschijnlijk vooral aankomen op informatiecultuur (zie hierboven). Daarnaast zijn er echter nog wel dingen mogelijk. Het is vooral belangrijk om bij veranderingen in het project, bijvoorbeeld als de doelen veranderen, als er nieuwe mensen aan gaan werken of er nieuwe applicaties in gebruik worden genomen, informatiebeheer in het oog te houden. Projectleiders en andere leidinggevendenden moeten erop getraind zijn om in dat soort situaties, waarin zaken snel over het hoofd gezien kunnen worden, advies te blijven zoeken om te voorkomen dat de veranderingen het informatiebeheer aantasten.

Een vraag die tot nu toe onbeantwoord is gebleven, is wat er dan precies bewaard en geregistreerd zou moeten worden. Het exacte antwoord hierop zal per project verschillen, maar voor een volledige dossiervorming zijn een aantal zaken in ieder geval van belang. Ten eerste is het overduidelijke antwoord natuurlijk de broncode. In dit stuk is de aandacht vooral uitgegaan naar de beperkingen van de broncode, maar dat wil niet zeggen dat het nutteloze informatie is. Zo maakt het bewaren van broncode gedetailleerde inspectie door experts mogelijk, en is het daarmee essentieel voor transparantie en uitlegbaarheid, maar kan het bijvoorbeeld ook samenwerking tussen gemeenten faciliteren. Broncode zou niet alleen bewaard moeten worden, maar actief openbaar moeten worden gemaakt. Hiervoor zou een landelijk register in het leven kunnen worden geroepen, waarin gemeenten en andere overheidsorganisaties broncodes uploaden met een korte toelichting over de werking van het algoritme, welke data het verwerkt en wat er met de output van het algoritme wordt gedaan.

De in- en outputdata zouden ook bewaard moeten worden, maar hierbij spelen privacyoverwegingen een belangrijke rol. De kans is namelijk groot dat hier gevoelige persoonsgegevens tussen zitten, die ervoor kunnen zorgen dat bewaren geen wenselijke optie is. In het geval van inputdata zou dan in ieder geval moeten worden bijgehouden welke soorten gegevens er verwerkt zijn, zodat op basis hiervan mogelijke ongewenst discriminerende patronen kunnen worden aangetoond. Aangezien outputdata meestal de basis is voor een beslissing – bijvoorbeeld of iemand verder onderzocht moet worden op bijstandsfraude – moet transparantie hierbij zwaarder wegen dan bij input. Een burger over wie een beslissing is genomen door of met behulp van een algoritme, zou hierover proactief geïnformeerd moeten worden door de overheidsorganisatie, om het bewustzijn van algoritmegebruik te verhogen. Naast de beslissing zou ook de onderbouwing ervan – in het kader van uitlegbaarheid – gecommuniceerd moeten worden naar de betreffende burger. Naast het bewaren van de outputdata, is het dus ook belangrijk om vast te leggen en te bewaren wat er met de betreffende output gedaan is. Als het hele proces niet te reconstrueren valt, heeft alleen de ‘lege’ outputdata verder immers weinig nut.

Uitlegbaarheid in de praktijk

Het andere uitgangspunt voor de aanbevelingen is het borgen van uitlegbaarheid. Dit is een nieuw concept dat specifiek in het kader van algoritmegebruik is bedacht en verdient daarom extra aandacht in dit stuk, ook omdat het mogelijk een grote impact kan hebben op de omgang met algoritmen. De term is in een Nederlandse context vooral bekend vanuit de richtlijnen van het ministerie van Justitie en Veiligheid, zoals besproken in hoofdstuk 2. Hier speelt het vooral een rol in het opvangen van de tekortkomingen van transparantie, door verdergaande openheid te eisen in de vorm van uitlegbaarheid. Het gaat hier ook nadrukkelijk om uitlegbaarheid aan burgers die te maken kunnen krijgen met processen waarbij een algoritme een rol speelt, en die impact op hen hebben. De Europese Commissie benadrukt eveneens dat uitlegbaarheid betrekking heeft op burgers en een belangrijke component is van hun concept van *human-centric AI*.¹⁷⁰ Het is zo belangrijk omdat de werking van algoritmen voor de niet-specialist erg abstract is en de kennis erover nog onvoldoende breed in de maatschappij verspreid is, terwijl ze wel in belangrijke, impactvolle domeinen ingezet worden.

In de interviews is ook veel gesproken over de uitwerking van dit concept. Hieruit werd duidelijk dat bestaande processen en stukken waarschijnlijk niet voldoende zijn om aan de eis van uitlegbaarheid te voldoen. Voor zover er werk wordt vastgelegd en gedocumenteerd, is dit namelijk vooral voor intern gebruik, zodat latere collega's het kunnen teruglezen en begrijpen wat er gedaan is. Deze toelichtingen zijn dus niet toegespitst op de behoefte van de burger, die kwesties als algoritmische data-analyse hoogstens op het niveau van een geïnformeerde leek zal kunnen begrijpen. Over het algemeen zagen de geïnterviewden ook wel het belang van het uitlegbaar houden van algoritmegebruik in, maar men had wat meer moeite met het idee dat er aparte documentatie voor zou moeten worden opgesteld, omdat dit extra werk en nieuwe processen met zich meebrengt.

Hier is echter niet omheen te komen als het idee van uitlegbaar algoritmegebruik daadwerkelijk in de praktijk gebracht moet worden. Uitlegbaarheid naar de burger toe is immers een compleet andere informatiebehoefte dan het werk omschrijven voor hergebruik door collega's. Dit is ook een kwestie waarin archief- en informatiebeheer niet alleen een antwoord op kan geven, hiervoor is ook kennis van communicatie en voorlichting vereist, om de werking van het algoritme naar begrijpelijke taal om te zetten. Toch is hier ook een rol voor archief- en informatieprofessionals te vinden, namelijk in het ontsluiten van informatie. Archiefdiensten hebben hier veel ervaring mee en kunnen ondersteuning bieden bij het (actief) openbaar maken van documenten, zodat deze met de juiste context en informatie gepubliceerd kunnen worden, zodat duidelijk is aan welk project de

¹⁷⁰ Europese Commissie, *Building Trust*, 5.

documentatie gerelateerd is. Het toekennen van Persistent Identifiers, unieke en permanente identificatiecodes die veel voor digitaal erfgoed gebruikt worden, aan algoritmen kan ervoor zorgen dat bijvoorbeeld informatie op gemeentewebsites gekoppeld kan worden aan het landelijke register.

De inhoud van deze documentatie ten behoeve van uitlegbaarheid ligt redelijk voor de hand: de werking van het algoritme moet er duidelijk, in begrijpelijke taal, in uitgelegd staan. De doelgroep van deze documenten is de gemiddelde burger die met het algoritme in aanraking kan komen – bijvoorbeeld doordat gegevens van de betreffende persoon door het algoritme verwerkt worden. Om te zorgen dat de uitlegbaarheidsdocumentatie actueel blijft, zou elke significante wijziging in het project, die merkbare effecten heeft op de werking van het algoritme, vastgelegd moeten worden in deze documenten. Deze update zou snel na de invoering online gepubliceerd moeten kunnen worden, zeker wanneer het gaat om algoritmen die actief gebruikt worden. Het geldt echter ook voor experimenten die met algoritmen worden uitgevoerd, aangezien het risico bestaat dat het experiment overloopt in een permanent proces, waarbij anders waardevolle informatie verloren kan gaan.

Daarnaast moeten burgers, zoals eerder gezegd, actief geïnformeerd worden wanneer zij onderwerp zijn van een algoritmische analyse. In een brief of e-mailbericht zou moeten staan in het kader van welk project, met welk doel en door welk algoritme de analyse wordt uitgevoerd, en waar meer informatie over het algoritme kan worden gevonden – oftewel de online uitlegbaarheidsdocumentatie. Het doel hiervan is vooral om het bewustzijn van algoritmegebruik te vergroten, zodat burgers aangemoedigd worden hierover na te denken en goed geïnformeerd zijn over het algoritmegebruik van de overheid. Dit neemt een deel van de ‘transparantieparadox’ weg, wat weer een betere maatschappelijke discussie over het onderwerp mogelijk maakt.

Een laatste punt dat in dit kader nog aandacht verdient, is de samenwerking met commerciële partijen. Het grote probleem hier, is de spanning tussen transparantie en uitlegbaarheid aan de ene kant, en intellectueel eigendomsrecht aan de andere kant. Zodra een overeenkomst over gebruik van een commercieel ontwikkeld algoritme voor overheidsdoeleinden is gesloten, is het moeilijk om daar nog van af te wijken. Daarom is het van groot belang dat er duidelijke afspraken worden gemaakt bij de inkoop: het algoritme moet ten minste beschikbaar zijn voor controle door onafhankelijke experts en moet worden opgenomen in het landelijke register. Dit betekent niet per se dat de broncode volledig openbaar moet zijn, maar deze kan in ieder geval onder voorwaarden gecontroleerd worden. Verder moet de maker van het algoritme documentatie kunnen aanleveren die laat zien hoe het algoritme tot stand is gekomen en welke aannames hierbij zijn gemaakt, zodat de benodigde uitlegbaarheidsdocumentatie kan worden opgesteld. Deze documentatie moet dan terecht kunnen komen in een archiveringssysteem waarin het voor de lange

termijn bewaard kan worden. Het gaat er immers om dat het werk in de toekomst gecontroleerd kan worden. Daarvoor is goede, duurzame archivering van onmiskenbaar belang.

Aan het eind van dit hoofdstuk, is het goed om de aanbevelingen nog even op een rijtje te zetten. De eerste besproken aanbeveling, was het inrichten van een interdisciplinaire overlegstructuur over kwesties rondom algoritmegebruik. Dit is niet per se direct gerelateerd aan een van de vier genoemde criteria, of aan archivering, maar is meer een algemene aanbeveling om te zorgen dat er inzicht komt in alle verschillende facetten en belangen die bij een dergelijk project komen kijken. Ten tweede kwamen de verbetering van archiveringssystemen en de metadatering aan bod. Deze maatregelen komen de vindbaarheid en traceerbaarheid van stukken ten goede, en staan direct in verband met transparantie en verantwoording. Indirect dragen ze ook bij aan het borgen van eerlijkheid, omdat met behulp van de stukken gecontroleerd kan worden of er sprake is geweest van een eerlijke behandeling. Hetzelfde geldt eigenlijk voor het veranderen van de informatiecultuur, die vooral moet bijdragen aan volledigere dossiervorming.

Een belangrijke rode draad in het tweede deel van de aanbevelingen, is actieve openbaarmaking en proactieve voorlichting over die openbaarmaking. Het doel hiervan is om de tekortkomingen van 'gewone' transparantie op te vangen, zoals het probleem dat niet duidelijk is waar gepubliceerde informatie te vinden is, en dat de verantwoording voor het vinden hiervan volledig bij de burger komt te liggen, of dat alleen het openbaar maken van bestaande stukken onvoldoende informatie oplevert om de situatie begrijpelijk te maken. Hierin speelt uitlegbaarheid een belangrijke rol, door de verplichting te creëren dat het werk uitgelegd moet kunnen worden, wat gepaard zou moeten gaan met het opstellen van extra documentatie en proactieve voorlichting. Anders kan er geen sprake zijn van daadwerkelijke uitlegbaarheid naar de burger toe, en dit vormt een goede aanvulling op transparantie- en verantwoordingsversterkende maatregelen als verplichte opname in een algoritmeregister en het bewaren van werkdocumentatie. Met deze combinatie van maatregelen, kunnen gemeenten grote stappen maken in het borgen van transparantie, verantwoording en uitlegbaarheid bij algoritmegebruik, en indirect ook bijdragen tot eerlijker algoritmegebruik.

De aanbevelingen opgesomd

- Creëer binnen de gemeente een interdisciplinaire overlegstructuur voor kwesties rondom algoritmegebruik.
- Verbeter de archiveringssytemen en de metadatering door een *archiving-by-design* perspectief te hanteren.
- Verander de informatiecultuur door deze gedetailleerder te onderzoeken en op strategisch niveau prioriteit te geven aan archivering en informatiebeheer, zodat de informatiecultuur gericht raakt op volledigheid en grondigheid.
- Zorg dat de beslissingen waarbij een algoritme betrokken is en de processen eromheen, volledig te reconstrueren zijn:
 - Implementeer *archiving-by-design* door vanaf een vroeg projectstadium actief advies in te winnen van archief- en informatiespecialisten rondom duurzame archivering en dit te blijven doen.
 - Bewaar de broncode en voeg deze toe aan een landelijk register.
 - Bewaar de in- en outputdata als onderdeel van het werkproces waarbij het algoritme ingezet is.
- Neem maatregelen om uitlegbaarheid te borgen:
 - Verhoog het maatschappelijk bewustzijn en faciliteer het debat door burgers actief en volledig te informeren over gebruik van algoritmen op vlakken die hen raken.
 - Stel nieuwe documentatie op waarin de werking van het algoritme helder wordt uitgelegd, maak deze duidelijk beschikbaar en koppel deze aan het landelijk register.
 - Maak duidelijke afspraken rondom uitlegbaarheid en transparantie met commerciële partners bij de inkoop van diensten waar algoritmen bij betrokken zijn.

Conclusie

Bij elk aspect van algoritmegebruik is een ding altijd duidelijk: er is nog geen definitieve ‘standaard’ en we zitten nog middenin de ontwikkelingen. Dat zal ongetwijfeld nog wel even zo blijven, met de verwachte groei en de toenemende geldstromen in het veld van de kunstmatige intelligentie. De eerste neiging is dan wellicht om te wachten met maatregelen, zodat de technologie alle ruimte krijgt om zich te ontwikkelen en er geen beleid wordt gemaakt dat in korte tijd alweer achterhaald is. In het geval van algoritmen is het echter niet langer verantwoord om te wachten, nu deze technologie een steeds duidelijkere plek krijgt in het functioneren van Nederlandse overheden en een rol speelt bij het nemen van beslissingen die burgers kunnen raken. De SyRI-rechtszaak heeft al laten zien dat algoritmen die in zo’n situatie gebruikt worden niet altijd gerechtelijke toetsing kunnen doorstaan, onder meer wegens een gebrek aan transparantie en een onevenredig grote inbreuk op de privacy door de koppeling van datasets.

Het is dus evident dat overheden ter verantwoording geroepen moeten kunnen worden voor hun algoritmegebruik. Het archief is altijd een van de belangrijkste instrumenten van bureaucratische overheden geweest om die verantwoording te kunnen afleggen: het op papier vastgelegde overheidshandelen was daar terug te vinden, waardoor de sporen van de overheid gevolgd konden worden. Inmiddels kunnen we in ieder geval niet meer spreken van ‘op papier vastgelegd’. Het archief van een digitaal werkende organisatie bestaat uit elektronische gegevens, volgens bepaalde indelingen weergegeven en opgeslagen op tal van opslagmedia, die ook lang niet altijd fysiek aanwezig zijn op de plek waar gewerkt wordt dankzij cloudopslag. Deze scriptie is ook op meerdere apparaten geschreven, met behulp van een synchronisatiedienst die de bestanden zowel lokale harde schijven als op hun eigen servers bewaart. Het handelen van een digitaal werkende overheid dient echter net zo goed te reconstrueren te zijn, zeker wanneer het gaat om de complexiteit die algoritmen inmiddels hebben bereikt.

Welke rol ligt hierin voor archivering? Dat was in wezen de centrale vraag voor dit onderzoek: hoe kunnen documentatie en archivering bijdragen aan verantwoord, eerlijk, transparant en uitlegbaar algoritmegebruik door Nederlandse gemeenten? Om een antwoord op deze vraag te kunnen geven, moest eerst de context in kaart gebracht worden. Dit begon in hoofdstuk 1 in de meest brede zin, namelijk de verhouding tussen technologieën en de politiek-maatschappelijke contexten waarin ze zich bevinden. Op basis van het werk van Verbeek is er een duidelijke beginpositie mogelijk: mens, maatschappij en techniek zijn onlosmakelijk met elkaar verbonden en zijn bepalend voor elkaars definities en functioneren. Het is voor de mens onmogelijk om niet beïnvloed te worden door techniek, maar de mens heeft wel de mogelijkheid om zelf de vorm en mate van beïnvloeding te bepalen.

In het geval van algoritmen was ook heel duidelijk te zien dat deze in alles verweven zijn met de context waarin zij worden gemaakt en gebruikt, waarbij sprake is van sterke wederzijdse beïnvloeding. Zo lieten Dana Boyd en Kate Crawford zien hoe bronnen die als ‘ruwe’, primaire data werden beschouwd, gevormd en beïnvloed waren door sociaal-culturele factoren, wat weer invloed heeft op de werking van algoritmen die met zulke datasets zouden werken. Claudia Aradau en Tobias Blanke toonden de andere zijde van de medaille, door te beargumenteren dat de manier van filteren en sorteren die algoritmen toepasten invloed had op het denken over veiligheid en dreiging. John Danaher wees op de politieke risico's van onvoldoende inzichtelijk algoritmegebruik, door het te benaderen vanuit een legitimiteitsperspectief. Archiefdeskundigen Jeurgens, Glaudemans en Verburgt benadrukten tenslotte het belang van een digitale wending in het archiefwezen, met een bredere blik op wat bewaard en vastgelegd zou moeten worden, om te zorgen dat het archief de rol van verantwoordingsinstrument kon blijven vervullen.

Uit het onderzoek naar de bestaande juridische- en beleidskaders bleek dat er zeker kaders bestaan rondom eerlijkheid, transparantie en verantwoording. Het gaat dan met name om de Archiefwet, de WOO en de uitvoeringswet AVG. De eerste twee hebben vooral betrekking op het borgen van transparantie en verantwoording, om de overheid controleerbaar te houden, terwijl de AVG actieve transparantie over gegevensgebruik en eerlijke gegevensverwerking afdwingt. Voor het specifieke geval van algoritmen zijn de richtlijnen van het ministerie van Justitie en Veiligheid en de Europese beleidsstukken het meest relevant. Zowel het ministerie als de EU benadrukken het belang van eerlijk en uitlegbaar algoritmegebruik, waarbij zij uitlegbaarheid zien als een verlengstuk van transparantie. Zelfs met maximale transparantie zou de werking van veel algoritmen namelijk nog steeds onbegrijpelijk blijven voor burgers die ermee in aanraking komen. Een rode draad door al deze stukken is verder dat ze het belang van archivering en documentatie benoemen, maar er – met uitzondering van de Archiefwet – weinig concrete eisen aan stellen. Zelfs de vernieuwde Archiefwet blijft hier in gebreke, omdat er onvoldoende aandacht is voor het unieke karakter van digitale informatie en voor de nieuwe structuren en contexten die de digitalisering met zich mee heeft gebracht.

De interviews boden een venster in de praktijk van zowel het gemeentelijk archief- en informatiebeheer als het werken met algoritmen. Uit gesprekken met beide groepen geïnterviewden bleek dat archivering veelal geen prioriteit is bij gemeenten: er wordt nog veel over gedacht als iets dat ‘achteraf’ kan, waar geen urgentie achter zit. Enkele geïnterviewden gaven ook aan dat ze zagen dat dit leidde tot incomplete dossiervorming, waarbij belangrijke delen van het proces of zelfs het eindproduct – bijvoorbeeld een vergunning – niet meer te traceren waren. Op het gebied van algoritmen was de voornaamste rode draad dat velen nog in de testfase zitten en dat er veel samenwerkingen zijn met commerciële partijen, met als doel de efficiëntie van een specifiek

werkproces te verhogen. Hierbij denkt men zeker na over uitlegbaarheid en verantwoording, maar het verband met archivering en documentatie wordt maar zelden gelegd.

De praktische risico's hiervan zijn groot. Experimenten kunnen immers geleidelijk vaste werkprocessen worden, die dan met onvoldoende oog voor archivering worden geïmplementeerd. Dit kan ervoor zorgen dat er tal van beslissingen worden genomen, met algoritmische ondersteuning, die niet meer te traceren en uit te leggen zijn, omdat alle, of sommige, 'sporen' ervan ontbreken. Met name bij complexe processen als algoritmen, die omringd worden door contexten van andere werkprocessen, technologische infrastructuren en culturele omstandigheden, kan het verlies van een klein deel van de context er al voor zorgen dat het overheidshandelen niet meer goed gereconstrueerd kan worden. Zo wordt het voor overheden onmogelijk om hun handelen uit te leggen aan burgers en kunnen burgers, misschien nog wel belangrijker, overheden niet meer goed ter verantwoording roepen. De hoofdvraag zoals hierboven geformuleerd, is dan ook eigenlijk heel kort te beantwoorden. Archiverings- en informatiebeleid levert een cruciale en onmisbare bijdrage aan het borgen van transparantie, verantwoording, uitlegbaarheid en eerlijkheid.

De eerste drie zijn duidelijk aan bod gekomen in de voorgaande hoofdstukken, maar eerlijkheid is nooit helemaal direct besproken. Dat komt ook doordat archivering eerder een indirecte rol speelt in de borging van eerlijkheid dan een directe. Eerlijkheid in algoritmen zal vooral moeten komen uit de omgang met biases in de programmatuur, de gebruikte data en in de omgang met de uitkomsten van het algoritme in het werkproces als geheel. Archivering speelt hier geen directe rol in. Het is echter wel belangrijk, omdat met behulp van de archieven kan worden gereconstrueerd of er sprake is geweest van oneerlijk handelen, bijvoorbeeld in de vorm van ongewenste discriminatie. Verder kunnen documentatie- en archiveringsprocessen ook bijdragen aan bewustwording van eerlijkheid: als er meer keuzes expliciet vastgelegd en bewaard moeten worden, zorgt dit er ook voor dat er meer over nagedacht wordt.

Transparantie en uitlegbaarheid liggen in zekere zin in elkaars verlengde. Strikt genomen is goede archivering natuurlijk geen randvoorwaarde voor transparantie – dat vereist alleen dat stukken openbaar gemaakt worden. Voor kwalitatief goede, zin- en betekenisvolle transparantie is het echter wel een duidelijke voorwaarde, anders wordt er onvolledige en mogelijk misleidende informatie verspreid. Uitlegbaarheid is wat complexer. Dit heeft namelijk ook betrekking op de technologie die in theorie gebruikt kan worden: als deze onmogelijk uitgelegd kan worden, moet er al van het gebruik worden afgezien. Als er voor een algoritme is gekozen dat uitgelegd kan worden, gaat archivering wel weer een rol spelen in het uitleggen van het werk. Op basis van documentatie kan het werk immers worden uitgelegd en goede archivering is nodig om te zorgen dat de voor de

uitleg benodigde documentatie op de juiste manier en met voldoende context gepubliceerd kan worden.

In zekere zin komt in verantwoording alles samen. Dit komt vooral omdat het om archivering gaat, dat verantwoording altijd als een van de belangrijkste doelen heeft gehad. De reden dat transparantie en uitlegbaarheid zo belangrijk zijn, is dat ze het handelen van de overheid zichtbaar maken voor burgers. Dit maakt democratische controle mogelijk, waarmee vervolgens verantwoording afgedwongen kan worden. Dat is ook de reden dat goede archivering en goed informatiebeheer een zodanig grote rol spelen in de aanbevelingen uit hoofdstuk 4. De kern hiervoor ligt in de informatiecultuur: zolang het belang van goed informatiebeheer en volledige documentatie niet doordringt bij medewerkers, zal overheidsinformatie altijd gebrekkig blijven. Het belang hiervan zal vooral ook op strategisch niveau moeten doordringen, zodat het een prioriteit kan worden binnen overheidsorganisaties en het idee genormaliseerd kan worden dat kleine werkzaamheden rondom informatiebeheer – zoals het op de juiste wijze opslaan en metadateren van stukken – bij de dagelijkse werkzaamheden horen. Hopelijk kan dit ook gepaard gaan met de ontwikkeling van betere systemen, die de last die dit werk ontegenzeggelijk met zich meebrengt kunnen verlichten.

Er ligt echter ook een belangrijke taak voor de archiefgemeenschap. Deze moet zich proactief opstellen en laten zien dat zij de kennis in huis heeft om te helpen bij deze vraagstukken. Met name bij algoritmen, die nu vooral als experiment worden gezien en waarbij veel nieuwe problematiek naar boven komt, liggen kansen. Men is op zoek naar een verantwoorde manier om deze technologie te implementeren en het is een technologie die zich in theorie uitstekend leent voor input van archief- en informatieprofessionals. Algoritmen draaien immers op gegevens: ze verwerken bestaande gegevens en genereren hier nieuwe gegevens mee. Goed informatiebeheer is dus cruciaal: als datasets niet meer terug te vinden zijn, kunnen ze nooit goed als input gebruikt worden. Verder heeft de SyRI-zaak wel duidelijk gemaakt dat fouten in de omgang met informatie grote gevolgen kunnen hebben en dat overheden verantwoordingen moeten kunnen afleggen over hun algoritmegebruik. Hiervoor is goed archief- en informatiebeheer wederom van cruciaal belang. De archiefsector moeten echter wel kunnen laten zien dat ze die kennis ook daadwerkelijk in huis heeft en zal ook het initiatief moeten nemen om te proberen een gesprekspartner te worden rondom algoritmen, anders zullen deze kansen waarschijnlijk blijven liggen.

De video van Oliver Thorn waarmee de inleiding van deze scriptie opende, volgt lange tijd een vast patroon. Telkens als de kritische clubbezoeker de redenering van de uitsmijter onderuit weet te halen en er geen uitweg meer mogelijk lijkt, verschijnt er een terugspoelteken in beeld, waarna de uitsmijter een andere redenering probeert. Zo duurt de socratische dialoog een tijd voort, totdat de ware problematiek naar voren komt: de uitsmijter wil de tijd niet nemen om na te denken over de

werking van de technologie die hij hanteert en de maatschappelijke gevolgen van het geheel: 'I don't wanna think about what this machine does. I wanna get through my shift, and go home, and on the weekend I wanna get drunk and watch Celtic beat the Rangers.' Een houding die we hem moeilijk kwalijk kunnen nemen, en waar waarschijnlijk iedereen zich wel deels in zal herkennen, ook degenen die er andere weekendactiviteiten op nahouden.

Na een laatste 'I don't want to think', gevolgd door een ongemakkelijke blik van de clubbezoeker, begint de aftiteling – het spel is afgelopen. Thorn legt hiermee op pijnlijk duidelijke wijze de kern van de problematiek rondom algoritmegebruik bloot: het liefst willen we niet teveel nadenken en stilstaan bij de grotere gevolgen van kleine dagelijkse handelingen, zeker als het om technologie gaat. We kunnen er echter niet omheen, zeker niet in een tijdperk waarin zulke grote technologische ontwikkelingen plaatsvinden met de potentie om enorme gevolgen te hebben voor de manier waarop we onze persoonlijke levens en ons openbaar bestuur inrichten. In lijn met het denken van Verbeek is het duidelijk dat we mens, maatschappij en algoritme niet los van elkaar kunnen zien en dat algoritmen waarschijnlijk niet zomaar weg zullen gaan: beïnvloeding is onvermijdelijk. We hebben echter ook de mogelijkheid en zelfs de verplichting om bewust na te denken over welke beïnvloeding wenselijk is. Deze scriptie is een eerste poging om een raamwerk op te zetten om deze beïnvloeding op een manier in te richten die past bij een democratische rechtsstaat, maar voor een echte oplossing is meer nodig dan alleen een enkele masterscriptie. Het vraagt om breed, langdurig, multidisciplinair en misschien wel eindeloos maatschappelijk debat. De taak voor burgers, academici, politici, bestuurders en ambtenaren voor de komende jaren zal worden om dat debat te voeren en verder te werken aan een wenselijke implementatie van deze technologie. Alleen zo kan er ooit sprake zijn van echt eerlijke algoritmen.

Bijlage 1: Literatuurlijst

- Alpaydin, Ethem, *Introduction to Machine Learning* (Cambridge MA 2010).
- Ananny, M., & Crawford, K., 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability', in: *New Media & Society* 20:2 (2018), 973–989.
- Andresen, Herbjørn, 'A discussion frame for explaining records that are based on algorithmic output', in: *Records Management Journal* 30:2 (2020), <https://doi.org/10.1108/RMJ-04-2019-0019>.
- Aradau, C., & Blanke, T., 'Governing others: Anomaly and the algorithmic subject of security.' *European Journal of International Security*, 3:1 (2018) 1-21.
- Aradau, C., Blanke, T., Kaufmann, M., & Jeandesboz, J., 'Politics of prediction: Security and the time/space of governmentality in the age of big data.' *European Journal of Social Theory*, 20:3 (2017) 373–391. <https://doi.org/10.1177/1368431016667623>.
- Bantwal Rao, M., Jongerden, J., Lemmens, P., & Ruivenkamp, G., 'Technological Mediation and Power: Postphenomenology, Critical Theory, and Autonomist Marxism.', in: *Philosophy & Technology*, 28(3) (2015), 449–474. <https://doi.org/10.1007/s13347-015-0190-2>.
- boyd, danah & Kate Crawford, 'Critical Questions for BigData', *Information, Communication & Society*, 15:5 (2012), 662-679.
- Chandler, David, 'Digital Governance in the Antropocene: The Rise of the Correlational Machine', in: David Chandler en Christian Fuchs ed., *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data* (Westminster 2019), 23-42.
- Craglia, M. (Ed.) e.a., *Artificial Intelligence – A European Perspective* (Luxemburg 2018) 57-59. Te raadplegen op: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf>. Laatst geraadpleegd op 19 maart 2020.
- Crawford, K., & Ziewitz, M., 'Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics', in: *Science, Technology, & Human Values* 41: 1 (2016) 77–92.
- Danaher, John. (2016) 'The threat of algocracy: reality, resistance and accommodation'. *Philosophy & Technology*, vol. 29, pp. 245-268.
- Doove, S. & Otten, D. *Verkennd onderzoek naar het gebruik van algoritmen binnen overheidsorganisaties*. (Den Haag: CBS 2018). Te raadplegen op: <https://www.cbs.nl/nl-nl/maatwerk/2018/48/gebruik-van-algoritmen-door-overheidsorganisaties>.
- Driessen, Camil, 'Rechtbank: overheid moet stoppen met fraudesysteem', website *NRC Handelsblad*, 5 februari 2020. Te raadplegen op: <https://www.nrc.nl/nieuws/2020/02/05/rechtbank-overheid-moet-stoppen-met-fraudesysteem-a3989325>. Laatst geraadpleegd op 31 juli 2020.
- Erfgoedinspectie, Stadsarchief Rotterdam en Interbestuurlijk Toezicht Noord-Brabant, *Integrale casusaanpak door de RIECs: Duurzame toegankelijkheid van archieven in de informatieketen* (Den Haag 2019).

Europese Commissie, *Artificial Intelligence for Europe* (Brussel, 25 april 2018). Te raadplegen op: <https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF>.

Europese Commissie, *Assessing Trustworthy AI* (Brussel 2018). Te raadplegen op: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/2>.

Europese Commissie, *Building Trust in Human-Centric Artificial Intelligence* (Brussel, 8 april 2019). Te raadplegen op: <https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-168-F1-EN-MAIN-PART-1.PDF>.

Europese Commissie, *White Paper On Artificial Intelligence – A European approach to excellence and trust* (Brussel, 19 februari 2020), 1-2; 10. Te raadplegen op: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Glaudemans, Arnoud, en Jacco Verburgt, 'The Archival Transition from Analogue to Digital: Revisiting Derrida and Flusser', in: Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times* ('s-Gravenhage 2017) 120-137.

Greenfield, Adam, *Radical Technologies: The Design of Everyday Life* (New York 2017).

Hao, Karen en Jonathan Stray: 'Can you make AI fairer than a judge?', in: *MIT Technology Review*: <https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>, gepubliceerd 17 oktober 2019, laatst geraadpleegd op 4 januari 2020.

High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (Brussel 2019). Te raadplegen op: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

Hoofnagle, C. J., van der Sloot, B., & Zuiderveen Borgesius, F., 'The European Union General Data Protection Regulation: What It Is And What It Means,' in: *Information & Communications Technology Law*, 28:1 (2019), 65-98.

Introna, Lucas D., 'Algorithms, Governance, and Governmentality: On Governing Academic Writing' in *Science, Technology & Human Values* 41, (2016) 17-49.

Jeurgens, Charles, 'Threats of the data-flood. An accountability perspective in the era of ubiquitous computing', in : Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times* ('s-Gravenhage 2017) 196-211.

Jonge, Adriaan de, 'Rookgordijn' rondom fraudesysteem Nissewaard', in : *Binnenlands Bestuur*, 24 april 2020. Te raadplegen op: <https://www.binnenlandsbestuur.nl/sociaal/nieuws/rookgordijn-rondom-fraudesysteem-nissewaard.13026523.lynkx>. Laatst geraadpleegd op 27 juli 2020.

Kamerstukken II 2013/14, 33328

Ketelaar, Eric, *Foreword*, in: Frans Smit, Arnoud Glaudemans en Rienk Jonker ed, *Archives in Liquid Times* (Den Haag 2017), VII.

Kiran, A., & Verbeek, P., 'Trusting Our Selves to Technology.' *Knowledge, Technology & Policy*, 23(3) (2010), 409–427.

Kool, L., J. Timmer, L. Royakkers en R. van Est, *Opwaarderen- Borgen van publieke waarden in de digitale samenleving*. (Den Haag, Rathenau Instituut 2017).

Lemmens, Pieter, 'Social Autonomy and Heteronomy in the Age of ICT: The Digital Pharmakon and the (Dis)Empowerment of the General Intellect', in : *Foundations of Science* (2017) 287-296

Michetti, Giovanni, 'Provenance in the Archives: The Challenge of the Digital Environment', in: Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times* ('s-Gravenhage 2017) 228-247.

Milberry, Kate, 'The Art of Collectively Loving Well in the Digital Age', in: *Foundation of Science* (2017) 297-300

Ministerie van Justitie en Veiligheid, *Bijlage bij Brief over waarborgen tegen risico's van data-analyses door de overheid* (Den Haag 2019). Te raadplegen op:

<https://www.rijksoverheid.nl/documenten/rapporten/2019/10/08/tk-bijlage-over-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid>.

Ministerie van Onderwijs, Cultuur en Wetenschap, *Van Archiefwet 1995 naar Archiefwet 202x: de belangrijkste wijzigingen van de nieuwe Archiefwet 202x, met korte toelichting*. (Den Haag 2019), te raadplegen op: <https://www.internetconsultatie.nl/archiefwet/document/5165>.

Ministerie van Welzijn, Volksgezondheid en Cultuur, *Memorie van Toelichting Archiefwet 1995* (Den Haag 1995), te raadplegen op: <https://www.inspectie-oe.nl/publicaties/publicatie/2014/11/12/memorie-van-toelichting-archiefwet-1995>.

Nationaal Archief, *Archiveren by design*, website Nationaal Archief: <https://www.nationaalarchief.nl/archiveren/kennisbank/archiveren-by-design>. Laatst gewijzigd op: onbekend. Laatst geraadpleegd op 7 mei 2020.

O'Neill, Cathy, *Weapons of Math Destruction. How big data increases inequality and threatens democracy* (Londen 2017).

Oliver, Gillian en Fiorella Foscari, *Records Management and Information Culture : Tackling the People Problem* (London 2014).

Otterlo, Martijn van (2017) 'From intended archivists to intentional algivists. Ethical codes for humans and machines in archives' in Frans Smit, Arnoud Glaudemans and Rienk Jonker (eds) *Archives in Liquid Times*, ('s-Gravenhage 2017), 266-293.

Philosophy Tube, *Data*. Gemaakt door Oliver Thorn. 31 januari 2020, te raadplegen op: <https://www.youtube.com/watch?v=fCUTX1jurJ4>. Laatst geraadpleegd op 31 juli 2020.

Pybus, Jennifer, Mark Coté, and Tobias Blanke. "Hacking the Social Life of Big Data." *Big Data & Society*, (December 2015). doi:[10.1177/2053951715616649](https://doi.org/10.1177/2053951715616649).

Reigeluth, Tyler (2014) 'Why data is not enough: digital traces as control of self and self-control' in *Surveillance & Society* 12(2) (2014) 243-254.

Røssaak, Eivind (2016) 'Memory and Media. Archival tasks in the age of algorithms' in *Schetsboek digitale onderzoekomgeving en dienstverlening* Den Haag, 32-41
<https://docplayer.nl/19659253-Schetsboek-digitale-onderzoekomgeving-dienstverlening-van-vraag-naar-experiment.html> (Links to an external site.)

Totta Data Lab, *Technische documentatie voorspelmodel bijstandsfraude Gemeente Nissewaard* (2019). Te raadplegen op:
https://nissewaard.notubiz.nl/document/8567516/1/Beantwoording_LOB%2C_SyRi-uitspraak_voor_algoritme_Totta_Data_Lab%2C_bijlage. Laatst geraadpleegd op 27 juli 2020.

Upward, Frank, Barbara Reed, Gillian Oliver en Joanne Evans, *Recordkeeping Informatics for a Networked Age* (Clayton 2017) Hoofdstuk 8, te raadplegen op:
http://books.publishing.monash.edu/apps/bookworm/view/Recordkeeping+Informatics+for+a+Networked+Age/207/19_chapter-title-8.html.

Van Est, R. & J.B.A. Gerritsen, with the assistance of L. Kool, *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality—Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)* (Den Haag, Rathenau Instituut, 2017).

Veenstra, Anne Fleur van e.a., *Quick scan AI in de publieke dienstverlening* (TNO, 2019). Te raadplegen op:
<https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/04/08/quick-scan-in-de-publieke-dienstverlening/TNO+-+AI+in+de+overheid+eindpresentatie+08042019.pdf>. Laatst geraadpleegd op 29 juli 2020.

Verbeek, P., 'Resistance Is Futile: Toward a Non-Modern Democratization of Technology.' *Techné: Research in Philosophy and Technology*, 17:1 (2013), 72–92.
<https://doi.org/10.5840/techne20131715>.

Verbeek, P., *Op de vleugels van Icarus : hoe techniek en moraal met elkaar meebewegen* (Rotterdam 2014).

Verbeek, P. The Struggle for Technology: Towards a Realistic Political Theory of Technology. *Foundations of Science*, 22:2 (2017), 301–304.

Verbeek, Peter Paul, 'Designing the Public Sphere: Information Technologies and the Politics of Mediation', p. 217-227. In: Luciano Floridi (ed), *The Onlife Manifesto. Being Human in a Hyperconnected Era* (Heidelberg-New York-Dordrecht-London 2015). Open Access:
<https://www.springer.com/la/book/9783319040929> .

Verbeek, P.P. (2016). 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research'. In: J.K. Berg O. Friis and Robert C. Crease, *Technoscience and Postphenomenology: The Manhattan Papers* (Londen 2016), 189-204.

VNG Realisatie, *Impactanalyse Richtlijnen voor het toepassen van algoritmes door gemeenten* (Den Haag 2020).

Wetenschappelijke Raad voor het Regeringsbeleid (2016). *Big Data in een vrije en veilige samenleving* (Amsterdam 2016).

Yeo, Geoffrey, 'Information, records, and the philosophy of speech acts' in Frans Smit, Arnoud GlauDEMans and Rienk Jonker (eds) *Archives in Liquid Times* ('s-Gravenhage 2017) 93-118.

Zuiderveen Borgesius, Frederik, *Discrimination, artificial intelligence, and algorithmic decision-making* (Strasbourg 2018). Te raadplegen op: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>.

Bijlage 2: *interview guides*

Interview guide medewerkers algoritmeprojecten

Wat wil ik weten?

- Welke werkprocessen zijn er te onderscheiden in AI-projecten bij gemeenten?
- Hoe wordt het werk nu gedocumenteerd? Waarom/voor wie?
- Hoe wordt er invulling gegeven aan eerlijkheid transparantie, verantwoordelijkheid en uitlegbaarheid?

Introductie

- Voorstellen: wie ben je, wat houdt je functie in?
- Hoe ben je betrokken bij projecten met algoritmen?
- Kun je een voorbeeld noemen van een project?

Werkprocessen

- In welke fases verloopt het project? **[evt vragen met voorbeeld omschrijven]**
- Welke werkprocessen zijn daarin terug te vinden?
 - Let hierbij op dat het concreet blijft
- Hoe verloopt dit in de praktijk?

Documentatie

- Welke maatregelen nemen jullie om goede documentatie te borgen?
- Zijn er informatie- of archiefspecialisten betrokken bij de projecten?
- Zijn jullie tevreden met de informatie- en documentatiesystemen die je tot je beschikking hebt?
 - Waarom (niet?)
- Met welk doel documenteren jullie het werk?

Waarden

- Zijn er discussies op de werkvloer over eerlijkheid, transparantie, uitlegbaarheid en verantwoordelijkheid?
 - Zo ja: waar komen deze discussies uit voort?
 - Zo nee: zijn er andere manieren waarop deze thema's aan bod komen?
- Houden jullie rekening met deze waarden in het opzetten van het project?
 - Zo nee: komt het terug in andere fasen?
- Hebben deze waarden weleens in de weg gestaan bij een project?
 - Hoe is daar toen mee omgegaan?
- Werken jullie weleens samen met commerciële partijen?
 - Hoe wordt daar omgegaan met eigendom v. data of algoritmen?

Einde

Ruimte maken om zelf wat te zeggen, toevoegingen etc.

Interview guide archief- en informatieprofessionals

Wat wil ik weten?

- Hoe wordt gemeentewerk gearchiveerd?
- Welke obstakels zijn er hierbij?
- Hoe worden de regels en procedures rondom archivering gehandhaafd?
- Hoe wordt doorgaans binnen gemeenten over archivering gedacht?
- Welke plaats heeft archivering in werkprocessen?
- Zijn archief- en informatieprofessionals betrokken bij projecten met algoritmen?

Introductie

- Voorstellen: wie ben je, wat houdt je functie in?
- Doorvragen over functie, vragen naar voorbeelden van processen of activiteiten

Documentatieprocessen bij gemeentes

- Wat voor soort informatiesystemen zijn doorgaans in gebruik bij gemeenten?
- Welke plaats neemt archivering doorgaans in binnen werkprocessen?
- Hoe wordt deze informatie overgebracht aan archiefinstellingen?
- Welke regels en richtlijnen spelen hierbij een rol?
- Is het lastig deze te handhaven?

Handhaving en struikelblokken

- Hoe denken gemeenteambtenaren doorgaans over archivering?
- Heeft het een hoge prioriteit voor ze?
- Hoe ziet men archivering in verband met eerlijkheid en transparantie?
- Hoe is dit terug te zien in de informatie die overgedragen wordt?
- Wat zijn de grootste obstakels in het handhaven van archiveringsregels?

Betrokkenheid bij projecten met algoritmen

- Zijn archief- en informatieprofessionals betrokken bij projecten met algoritmen?
- Zo ja: in welke fase? Wat wordt ze gevraagd?
- Zo nee: waarom niet? Weleens geprobeerd om zelf actief een project op te zoeken? Hoe is dat gegaan?

Einde

Ruimte maken om zelf wat te zeggen, toevoegingen etc.