Hoe ‘algoprudentie’ kan bijdragen aan een verantwoorde inzet van machine learning-algoritmes

Meuwese, A.C.M.; Parie, J.; Voogt, A.

Hoe ‘algoprudentie’ kan bijdragen aan een verantwoorde inzet van machine learning-algoritmes

Anne Meuwese, Jurriaan Parie & Ariën Voogt[1]

Aan de hand van twee casusposities rond de inzet van machine learning-gedreven risicoprofilering door de gemeente Rotterdam en de gemeente Amsterdam wordt het begrip ‘algoprudentie’ geïntroduceerd en uitgewerkt. Deze nieuwe term verwijst naar concrete op casus gebaseerde en gedecentraliseerde oordeelsvorming over de verantwoorde inzet van algoritmes. Op grond van een analyse waaruit blijkt dat de algemene beginselen van behoorlijk bestuur ontoereikend zijn om deze algoritmes concreet te normeren, betogen de auteurs dat algoprudentie een nuttige aanvulling op en een concretisering van bestaande juridische kaders kan vormen.

1. Inleiding

Er gebeurt momenteel veel om de verantwoorde inzet van algoritmes in het publieke domein te borgen. Hoewel risicoprofilering en zwarte lijsten tot op de dag van vandaag een stempel op het imago van de Nederlandse overheid drukken, zijn er belangrijke stappen gezet. Denk aan het algoritmeregister, de Impact Assessment Mensenrechten en Algoritmes (IAMA), het Rijksbrede Algoritmekader en de nieuwe Directie Coördinatie Algoritmes (DCA) bij de Autoriteit Persoonsgegevens (AP). Toch leidt de ontwikkeling van dit soort kaderende instrumenten, net als organisatorische waarborgen, niet uit zichzelf tot verantwoorde algoritmes.[2] Ook de abstracte juridische normen die vanuit wetgeving en jurisprudentie[3] gelden voor de inzet van algoritmes door Nederlandse (overheids)organisaties, sorteren lang niet altijd het gewenste effect. Het uitblijven van het reguleringseffect is deels te verklaren vanuit de kloof tussen de algemene kaders en de concrete vraagstukken die spelen in de algoritmische praktijk.

In dit artikel betogen wij dat concrete op casus gebaseerde en gedecentraliseerde oordeelsvorming over de verantwoorde inzet van algoritmes een bijdrage kan leveren aan de nadere invulling van geldende juridische kaders. Het gaat hier om een beginnende praxis, die zich buiten het positieve recht voltrekt, maar daar indirect wel een bijdrage aan kan leveren. Wij noemen deze praktijk ‘algoprudentie’.

Hoewel algoprudentie ook buiten de publieke sector een rol kan spelen, illustreren wij de noodzaak en meerwaarde van algoprudentie in de context van machine learning-gedreven (ML) risicoprofilering door bestuursorganen. Daartoe bespreken wij allereerst in paragraaf 2 een juridisch kader dat op deze praktijk van toepassing is, in de vorm van drie relevante maar soms wat onderbelichte algemene beginselen van behoorlijk bestuur (abbb): het motiveringsbeginsel,[4] het zorgvuldigheidsbeginsel[5] en het beginsel van fair play.[6] Op grond van onze evaluatie dat deze beginselen tegelijkertijd relevant maar ook ontoereikend zijn om de algoritmische praktijk te normeren, betogen wij de noodzaak van contextualisering om tot concreet toepasbare normen te komen, en introduceren we ‘algoprudentie’ als mechanisme hiervoor.

Nieuwe vragen die opkomen vanuit de algoritmische praktijk worden niet voldoende geadresseerd door de abstracte en open normen in het bestaande juridische kader

Een van de eerste oordelen die wij tot de algoprudentie rekenen gaat over de inzet van ML-gedreven risicoprofilering bij bijstandsheronderzoek, zoals tot voor kort gehanteerd door de gemeente Rotterdam.[7] Deze casus, alsmede de hieraan verwante en inmiddels stopgezette 'Slimme check levensonderhoud' van de gemeente Amsterdam, behandelen wij in paragraaf 3 om te laten zien dat nieuwe vragen die opkomen vanuit de algoritmische praktijk niet voldoende worden geadresseerd door de abstracte en open normen in het bestaande juridische kader. Vervolgens demonstreren we in paragraaf 4 aan de hand van de casusbespreking welke concrete bijdrage algoprudentie kan leveren aan een verantwoorde inzet van algoritmes. Wij sluiten af met een meer algemene beschouwing over algoprudentie als nieuw concept in het juridische landschap (par. 5) en een conclusie (par. 6).

2. Evaluatie van de bestuursrechtelijke normering van ML-gedreven risicoprofilering

In deze paragraaf introduceren we de praktijk van machine learning-gedreven (ML) risicoprofilering door bestuursorganen en een normerend kader van drie algemene beginselen van behoorlijk bestuur die daarop van toepassing zijn. Wij betogen dat deze beginselen wel relevant zijn voor de algoritmische praktijk, maar contextualisering behoeven, waarvoor algoprudentie een veelbelovend mechanisme is.

2.1. Hoe werkt ML-gedreven risicoprofilering?

Het is bekend dat Nederlandse overheidsorganisaties gebruik maken van ML bij risicoprofilering. Zo kunnen gemeenten ML gebruiken om inwoners die een bijstandsuitkering ontvangen te selecteren voor heronderzoek.[8] In de analoge (niet-algoritmische) variant wordt door gemeenteambtenaren jaarlijks handmatig een profiel samengesteld om bijstandontvangers te selecteren voor heronderzoek (bijvoorbeeld 'alleenstaande mannen met een huisgenoot'), aangevuld met aselecte steekproeven en signaalgedreven selectie van inwoners. Naast deze analoge profileringsmethoden kan ML worden ingezet om door middel van data-analyse criteria op te stellen voor een risicoprofiel. Mits goed ingebed in risicobeheersmaatregelen (zoals organisatorische waarborgen, evaluatie- en documentatieverplichtingen) vergroot deze algoritmische werkwijze de effectiviteit en verkleint het de kans op discriminatie en willekeur – zo luidt althans de belofte. In de praktijk blijken deze beheersmaatregelen lastig te realiseren. Het is doorgaans een uitdaging om gebruikte ML gedegen te documenteren en te evalueren. En ook als de documentatie behoorlijk goed op orde is, blijft een rijtje lastige vragen over. Welke variabelen worden aan het algoritme gevoed en waarom? Van welk type ML-algoritme wordt gebruik gemaakt en waarom? En volgens welke prestatiemetrieken (waaronder metrieken om *fairness* en effectiviteit te meten) wordt het algoritme gemonitord en geëvalueerd? Dit zijn urgente vragen met een sterke normatieve lading waar geen pasklare antwoorden op bestaan. Zoals we in de casusbespreking illustreren, gaat het hier om waardenafwegingen die onlosmakelijk zijn verbonden aan technische aspecten van het modelleren van data. In de volgende sub-paragrafen analyseren we de rol die de abbb spelen bij het normeren van deze dimensie.

2.2. De abbb als ijkpunt voor ML-gedreven risicoprofilering

Als ML-gedreven risicoprofilering wordt ingezet zonder dat expliciet met publieke waarden rekening wordt gehouden, komt deze praktijk al snel in botsing met veel abbb, die deels zijn vastgelegd in de Algemene wet bestuursrecht (Awb).[9] Veel uitgangspunten van de abbb in hun huidige reikwijdte schuren met bepaalde kenmerken van ML-gedreven risicoprofilering: zoals enerzijds de aanname dat belanghebbenden in beginsel beschikken over alle voor bezwaar of beroep relevante informatie en anderzijds de inzet van complexe statistiek. Tegelijkertijd kan een verantwoorde inzet van ML mogelijk juist bijdragen aan het vervullen van verplichtingen die voortvloeien uit het zorgvuldigheidsbeginsel, aangezien datamodellen het informatiebeeld voor het bestuursorgaan (en dus potentieel ook voor de belanghebbende) completer kunnen maken.[10]

Veel uitgangspunten van de abbb schuren met bepaalde kenmerken van ML-gedreven risicoprofilering

Auteurs

1. Prof. dr. A.C.M. Meuwese is hoogleraar Public Law & Governance of Artificial Intelligence aan de Universiteit Leiden en lid van de raad van advies van Stichting Algorithm Audit. J.F. Parie MSc is wiskundige en A.W. Voogt MA is filosoof en promovendus aan de PThU en zij zijn beiden bestuurder van Stichting Algorithm Audit. De auteurs bedanken Ola Al Khatib, Robbert Bruggeman, Annemarie Drahmann en Joyce Esser voor hun waardevolle feedback op een eerdere versie, en Munish Ramlal voor de inspiratie voor de term 'algoprudentie'.

Noten

2. Zie bijvoorbeeld over visumverstrekking: nrc.nl/nieuws/2023/04/23/beslisambtenarenblijven-profileren-met-risicoscores-a4162837; over fraudecontrole door DUO en het UWV: www.nrc.nl/nieuws/2024/03/01/rapport-indirecte-discriminatie-bij-controles-op-fraude-met-uitwonendenbeurs-a4191795 en nos.nl/artikel/2482915-uwv-verzamelde-illegaal-gegevens-van-uitkeringsgerechtigden.

3. Rb. Den Haag, 5 februari 2020, ECLI:NL:RBDHA:2020:865 (*SyRI*).

4. Art. 3:46 en 3:47 Awb.

5. Art. 3:2 Awb.

6. Art. 2:4 Awb.

7. Over deze casus heeft een commissie van experts namens Stichting Algorithm Audit, waar twee van ons aan verbonden zijn als bestuurslid en een derde als lid van de raad van advies, onlangs een oordeel uitgebracht (Stichting Algorithm Audit, 'Risicoprofilering heronderzoek bijstandsuitkering' (AA:2023:02), bestaande uit een adviesrapport (AA:2023:02:A) en een probleemstelling (AA:2023:02:P).

8. Art. 53a en 64 Participatiewet bieden een rechtsgrond voor heronderzoek.

9. A.C.M. Meuwese, 'Digitalisation and administrative law: legal and administrative aspects', in: L. Jančová & M. Fernandes (red.), *Digitalisation and administrative law: European added value assessment*, Brussel: European Parliamentary Research Service (EPRS) 2022, Annex.I-VIII/1-48.

10. H.C.H. Hofmann, 'An introduction to Automated Decision Making (ADM) and Cyber-Delegation in the Scope of EU Public Law', INDIGO Working Paper, 21 juni 2021.

© akin bostanci / getty images

De consensus in de bestuursrechtelijke literatuur is dat uit de abbb inhoudelijke normen voor de praktijk van risicoprofilering door bestuursorganen voortvloeien.[11] Een belangrijke reden waarom naar de abbb gekeken wordt, is dat veel concrete Awb-normen alleen voor besluiten gelden, terwijl de inzet van een algoritme naar huidig recht geen 'besluit' in de zin van de Awb behelst. De abbb gelden daarentegen voor een breder scala aan handelingen, al staat de concrete codificatie van sommige beginselen wel in de sleutel van het besluitbegrip (zie bijvoorbeeld artikel 3:2 Awb). Dit laatste is echter geen bezwaar, aangezien het soort ML-toepassingen waar deze bijdrage over gaat, ook worden ingezet in een besluitvormingscontext. De rol die de abbb vervullen, of zouden moeten vervullen, als 'aanvullende gedragsnormen'[12] is interessant in het kader van het normeren van handelingen rondom de eigenlijke besluitvorming. Tegelijkertijd is de exacte betekenis van de abbb bij het reguleren van algoritmische besluitvorming vaak niet duidelijk en komt hun rol als 'gedragsnorm' niet sterk uit de verf. Zo is recent in dit tijdschrift nog de vraag gesteld of 'de beginselen van behoorlijk bestuur, zoals zorgvuldigheid en motivering (...) voldoende breed en flexibel [zijn] om inhoud te geven aan de roep om *explainability, accountability* en *transparency* van AI-gebruik'.[13] Van Ettekoven signaleerde al eerder dat in zaken die voor de bestuursrechter komen bestuursorganen moeite hebben om antwoord te geven op zelfs simpele vragen, zoals 'welke data gebruikt zijn of hoe een algoritme tot een besluit is gekomen' (dus laat staan de hierboven genoemde 'lastige vragen').[14] Het is dan ook niet voor niets dat de Afdeling advisering van de Raad van State bij algoritmische besluitvorming een verscherpte interpretatie van de beginselen van behoorlijk bestuur, en in het bijzonder het motiveringsbeginsel en het zorgvuldigheidsbeginsel voorstaat.[15]

2.3. Motiveringsbeginsel

Op basis van het motiveringsbeginsel moet voldoende duidelijk zijn op basis waarvan en waarom een bestuursorgaan een besluit neemt. Als zodanig heeft het motiveringsbeginsel een meta-karakter: het is dit beginsel dat de belanghebbende de gelegenheid geeft te weten te komen of in het kader van een bepaald besluit de andere beginselen ook zijn nageleefd. Het motiveringsbeginsel kan bij ML-gedreven risicoprofilering in het geding komen wanneer uitgelegd moet worden waarom een algoritme tot een bepaalde uitkomst is gekomen. Het is een uitdaging om op statistiek gebaseerde redeneringen om te zetten naar juridische redeneringen.[16] Bij onze casusbespreking van twee ML-algoritmes zullen wij ingaan op de (on)mogelijkheden rond het pleidooi om 'bestuursorganen [te] dwingen om de "berekeningen" door het algoritme in natuurlijke taal beschikbaar te stellen'[17] teneinde te voldoen aan het motiveringsbeginsel. Ook het College voor

Het is een uitdaging om op statistiek gebaseerde redeneringen om te zetten naar juridische

de Rechten van de Mens merkt enerzijds op dat het motiveringsbeginsel op dit moment al verplicht tot transparant algoritmegebruik, maar vraagt anderzijds om in het belang van de duidelijkheid en de achterliggende mensenrechtelijke belangen om een expliciete regeling in de Awb.[18] Wat wel en niet valt onder 'uitlegbaarheid' als niet-juridisch onderdeel van de juridische norm van het motiveringsbeginsel is volop in ontwikkeling, en concrete normen ontbreken dus vooralsnog.[19]

2.4. Zorgvuldigheidsbeginsel

Een minder prominente rol in de literatuur over 'digitaal bestuursrecht' is weggelegd voor het zorgvuldigheidsbeginsel. Dit is eigenlijk best vreemd: dit beginsel ziet immers op de totstandkoming van een besluit, en laat ML-gedreven risicoprofilering nu juist in die fase worden ingezet. Het verplicht overheidsorganisaties ertoe de omstandigheden te scheppen waarin een juist besluit kan worden genomen. Zo moeten relevante feiten en af te wegen belangen bekend zijn. Er moet een geschikte methode worden gebruikt om de belangen af te wegen en er dient een volledige belangenafweging plaats te vinden. Het zorgvuldigheidsbeginsel kan bij ML-gedreven risicoprofilering in het gedrang komen als de gebruikte inputdata incompleet of incorrect zijn en wanneer het risicoprofiel niet alle relevante feiten meeneemt. Maar zelfs als aan deze voorwaarden wordt voldaan, blijven er voor het zorgvuldigheidsbeginselen relevante vragen over: is ML bijvoorbeeld de 'geschikte' methode, en zo ja welke vorm van ML? Is de manier waarop een specifiek ML-model data en (persoons)kenmerken meeweegt wel juist? Ook hiervoor geldt dat het beginsel onvoldoende is ingevuld om concrete handvatten te bieden.

2.5. Fair play-beginsel

Het beginsel van *fair play*, ofwel correcte bejegening, dat als een verbod van vooringenomenheid van artikel 2:4 Awb deels gecodificeerd is, gaat over het zonder partijdigheid vervullen van taken door een bestuursorgaan. Het biedt het voordeel dat het ook buiten de specifieke besluitcontext van toepassing is. Vanwege de toespitsing op persoonlijke belangen in het tweede lid, in combinatie met het feit dat het beginsel geen grote rol speelt in de jurisprudentie,[20] wordt dit beginsel snel geassocieerd met ofwel de afwezigheid van corruptie ofwel onfatsoenlijke processtrategie aan de kant van het bestuursorgaan. In een digitaliseringscontext is er alle reden om dit beginsel ruimer te interpreteren, zoals Van Eck laat zien in haar proefschrift, waarin zij een 'welwillende houding' aan de kant van het bestuursorgaan en het niet 'willens en wetens onzorgvuldig handelen' tot de essentie bestempelt.[21] Het 'contextualiseren'[22] van de abbb, wat wij met Wolswinkel bepleiten, zou zich in het geval van dit beginsel goed kunnen toespitsen op nieuwe, digitale verschijningsvormen van vooringenomenheid. Vervolgens zou een daarop aansluitende inspanningsverplichting kunnen gelden om bij algoritmegebruik onwenselijke *bias* te voorkomen en eerlijkheid te waarborgen.

2.6. De noodzaak van contextualisering en de rol van algoprudentie

Bovenstaande analyse laat de relevantie van de abbb zien in de context van ML-gedreven risicoprofilering en wijst tegelijkertijd op de noodzaak van contextualisering om tot concreet toepasbare normen te komen. De contextualisering van bovengenoemde drie én andere beginselen, en de mogelijke herijking die daarop kan volgen, zullen zich niet vanzelf voltrekken. Tegelijkertijd blijkt het voor zowel de rechter, de wetgever als de toezichthouder lastig dit proces in gang te zetten. De rechter is zeker geïnteresseerd in het nader invullen van normen voor algoritmische besluitvorming, zoals de bekende jurisprudentie over SyRI en de Aerius-toepassing laat zien.[23] De rechter is echter afhankelijk van de (tot nu toe zeldzame) concrete zaken die voorkomen en in de zaken die er zijn, spelen de abbb een relatief kleine rol.[24] Bij de bestuursrechter speelt ook nog dat een oordeel over een algoritme altijd onderdeel zal zijn van een breder rechtmatigheidsoordeel en dus niet snel antwoorden zal kunnen bieden op alle concrete vragen die spelen in de ML-praktijk. Het is niet de taak van de wetgever om al te gedtailleerde normen te ontwerpen; zo laat de Europese AI-verordening de nadere invulling ook over aan geharmoniseerde normen. Het gegeven dat in het kader van de internetconsultatie van het wetsvoorstel versterking waarborgfunctie Awb geopteerd is voor een apart 'Reflectiedocument algoritmische besluitvorming en de Algemene wet bestuursrecht' laat zien dat ook de Awb-wetgever met het normeringsvraagstuk worstelt.[25] De toezicht-

11. Ch. Adriaansz, 'De rechtmatigheid van algoritmische besluitvorming in het licht van het zorgvuldigheidsbeginsel en het motiveringsbeginsel', *NTB* 2020/100; C.J. Wolswinkel, 'AR meets AI: een bestuursrechtelijk perspectief op een nieuwe generatie besluitvorming', *Computerrecht* 2020/4, p. 22-29; B.J. van Ettekoven & A.T. Marseille, 'Afscheid van de klassieke procedure in het bestuursrecht?', in: L.M. Coenraad e.a., *Afscheid van de klassieke procedure* (NJV-preadviezen), Deventer: Wolters Kluwer 2017, p. 260.
12. Wolswinkel 2020, p. 24.
13. B. J. van Ettekoven & J.E.J. Prins, 'Artificiële intelligentie en de Rechtspraak. Implicaties van de Europese AI Act en de noodzaak van een kompas voor toepassing en beoordeling van AI-systemen', *NJB* 2023/2633, afl. 36, p. 3156.
14. Algemene Rekenkamer, 'Rondetafelgesprek #1: discriminatie door algoritmes', rekenkamer.nl/onderwerpen/algoritmes/digitaal-event/rondetafelgesprek-1.
15. Advies Afdeling advisering Raad van State van 31 augustus 2018, *Kamerstukken II* 2017/18, 26643, nr. 557, p. 3.
16. Meuwese 2022.
17. Y.E. Schuurmans, A.E.M. Leijten & J.E. Esser, *Bestuursrecht op Maat* (in opdracht van BZK), Leiden: Universiteit Leiden 2020, p. 73, als bijlage aangeboden bij *Kamerstukken II* 2020/21, 29362, nr. 289.
18. College voor de Rechten van de Mens, 'Toetsingskader Discriminatie door Risicoprofielen', december 2021; College voor de Rechten van de Mens, 'In alle openheid: transparant algoritmegebruik door de overheid', position paper, juni 2023.
19. Wolswinkel 2020.
20. A.de Moor-van Vugt, 'Fair play – een vergeten beginsel', in: T. Barkhuysen, B. Marseille, W. den Ouden, H. Peters, & R. Schlössels (red.), *25 jaar Awb: In eenheid en verscheidenheid*, Wolters Kluwer, 2019, p. 393-403.
21. M. van Eck, *Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming* (diss. Tilburg), 2018, p. 28.
22. Wolswinkel 2020, p. 29.
23. Supra noot 3 en ABRvS 17 mei 2017, ECLI:NL:RVS:2017:1259 (*Aerius*); zie ook Wolswinkel 2020.
24. O.A. Al Khatib, 'De (on)zin van transparantie voor de bewijspositie van belanghebbenden bij algoritmische overheidsbesluitvorming', preadvies Jonge VAR 2023, [in definitieve versie nog te verschijnen].
25. internetconsultatie.nl/algoritmischebesluitvormingenawb/b1.

houder heeft in het verleden wel aangegeven aan normuitleg te gaan doen, maar tot op heden is hier weinig van gebleken. De institutionele impasse rond de concretisering van juridische normen voor de algoritmische praktijk die hieruit blijkt, moet worden doorbroken. Wij betogen dat algoprudentie hier een rol in kan spelen.

Deze bijdrage introduceert het begrip 'algoprudentie' voor concrete op casus gebaseerde en gedecentraliseerde oordeelsvorming over verantwoorde inzet van algoritmes. In de kern is algoprudentie een voortgaand gesprek tussen diverse partijen in de samenleving over de beslechting van normatieve vraagstukken die zich voordoen bij de inzet van algoritmische toepassingen, op basis van concrete oordelen over een casus. Algoprudentie kan bijdragen om op een transparante en effectieve wijze invulling te geven aan de abbb en andere juridische normen, en om de oordeelsvorming over algoritmische toepassingen te harmoniseren. Net als 'legisprudentie' en 'ombudsprudentie' wijkt het concept in belangrijke opzichten af van 'jurisprudentie'. Naast uiteraard het niet-bindende karakter, springen daarbij in het geval van algoprudentie het decentrale en niet-hiërarchische karakter in het oog. Het oordeel wordt niet overgelaten aan formele instanties, maar aan meer of minder formele maatschappelijke lichamen die dichter op de algoritmische praktijk staan. Om vooruitgang te kunnen boeken met de concretisering van abstracte juridische normen voor algoritmes zoals de abbb, is het namelijk essentieel dat ook zonder dat enige juridische procedure is gestart, uitspraken gedaan kunnen worden over voorliggende normatieve kwesties.

Tegelijkertijd is de aanspraak op het 'prudentiële' karakter van deze ontluikende praktijk essentieel. Als de organisaties en commissies die moeten besluiten over de normatieve aspecten van algoritmes deze zouden benaderen als kwesties die pragmatisch kunnen worden opgelost, wat slechts resulteert in '*best practices*', valt een essentieel element weg, namelijk een deliberatieve en gemotiveerde afweging waarin de *juistheid* expliciet wordt gethematiseerd. De concrete toepassing van beginselen als zorgvuldigheid en *fairness* heeft een interpretatieve gemeenschap nodig. In het geval van ML-gedreven risicoprofilering is deze ten eerste sterk interdisciplinair en ten tweede nog sterk in opbouw. De introductie van 'algoprudentie' kan daarom op het eerste gezicht enigszins voorbarig overkomen. Wij menen echter dat deze *framing* juist een positieve bijdrage kan leveren aan het professionaliseren en stroomlijnen van de inspanningen van deze gemeenschap.

Voordat wij de ontluikende praktijk van algoprudentie verder conceptueel uitwerken en juridisch inbedden, willen we aan de hand van casusonderzoek bovenstaande analyse van de abbb in de context van ML-gedreven risicoprofilering concretiseren, om vervolgens de potentie van algoprudentie te demonstreren.

De concrete toepassing van beginselen als zorgvuldigheid en *fairness* heeft een interpretatieve gemeenschap nodig

3. Casusonderzoek: de beginselen in de praktijk van ML-gedreven risicoprofilering

In deze paragraaf bespreken wij twee recente casussen van ML-gedreven risicoprofilering door Nederlandse overheden, die dienen ter illustratie van zowel de relevantie als de huidige onmacht van het zorgvuldigheidsbeginsel, het motiveringsbeginsel en het beginsel van *fair play* om grip te krijgen op ML-gedreven risicoprofilering in de praktijk.

3.1. Amsterdamse en Rotterdamse ML-gedreven risicoprofilering

Eén casus heeft betrekking op de gemeente Rotterdam tussen 2017 en 2021, de andere heeft recent gespeeld bij de gemeente Amsterdam. Zij verschillen op een aantal cruciale punten: het type algoritme dat werd ingezet, de fase in de uitvoering waarin zij een rol spelen en de wijze waarop het ontwikkelproces van het algoritme is gedocumenteerd. Vanwege deze verschillen geven de casussen een goed beeld van een evoluerende praktijk.

In 2022 introduceerde de gemeente Amsterdam de pilot 'Slimme check levensonderhoud'.[26] Dit algoritme kent een onderzoekwaardigheidsscore toe aan iedere nieuwe aanvraag van een bijstandsuitkering. De score wordt bepaald aan de hand van een *explainable boosting model* (ebm) ML-algoritme dat is getraind op aanvragen uit het verleden en bijbehorende data over de woonsituatie, eerdere bijstandsaanvragen, het inkomen en vermogen van de inwoner op het moment van aanvraag. Aanvragen die een score boven een bepaalde drempelwaarde kregen toegekend, worden door een medewerker nader onderzocht. Met de inzet van een dergelijk algoritme in de aanvraagfase, wil de gemeente Amsterdam de nadruk verleggen van controle achteraf naar een zo accuraat mogelijke toekenning. Na evaluatie van de pilot is begin 2024 besloten het algoritme niet in gebruik te nemen.[27]

Het bijzondere 'uitlegbare' (*explainable*) karakter van het ebm-algoritme komt voort uit de manier waarop de onderzoekswaardigheidsscore wordt bepaald. Het model berekent eerst aan de hand van complexe statistiek per kenmerk (woonsituatie, inkomen etc.) een score, waarna de scores bij elkaar worden opgeteld tot een eindscore. Dit generiek additieve karakter van de ebm-methode maakt dat van ieder kenmerk wordt bijgehouden wat de invloed is op de eindscore.

In Rotterdam is in de periode 2017-2021 een *extreme gradient boosting* (xgb)-algoritme gebruikt (en na controverse stopgezet) om aan de hand van risicoscores onrechtmatigheid te voorspellen voor burgers die al een bijstandsuitkering ontvangen. Dit xgb-algoritme is getraind om in meer dan 60 kenmerken van bijstandontvangers patronen te detecteren of iemand onrechtmatig een uitkering ontvangt. Aan de hand van de hoogte van de risicoscores zijn inwoners geselecteerd voor heronderzoek. Bij een xgb-algoritme is het complexer dan bij een ebm-algoritme om een verband te leggen tussen de overwogen kenmerken en de voorspelde risicoscore. De risicoscore wordt namelijk niet vastgesteld door scores per kenmerk bij elkaar op te tellen, maar wordt vastgesteld in een complexe kansberekening waar alle kenmerken worden

gecombineerd. Het is enkel in complexe statistische termen te achterhalen hoe een bepaald kenmerk heeft bijdragen aan de totstandkoming van een individuele risicoscore.[28] Dit *black box*-karakter geldt ook voor het ebm-algoritme, maar dan alleen voor de manier waarop een score per kenmerk wordt bepaald. Kortom, beide methoden zijn een *black box*, al heeft het xgb-algoritme een sterker *black box*-karakter dan het ebm-algoritme waardoor het minder uitlegbaar is.

3.2. Gemotiveerde, zorgvuldige en eerlijke ML-gedreven risicoprofilering

Hoe verhouden de gebruikte ebm- en xgb-profileringsmethodes zich tot het zorgvuldigheid-, motiverings- en fair play-beginsel? Het is van belang op te merken dat in beide casussen ML niet direct wordt ingezet voor het nemen van een besluit of een uitkering rechtmatig is, maar alleen voor de voorbereiding op een besluit, dat uiteindelijk door medewerkers wordt genomen na vervolgonderzoek. Uit de voorafgaande analyse blijkt echter dat de abbb nog steeds relevant zijn voor de gebruikte ML-methodes.

3.2.1. Motivering

Uit onze bespreking van het motiveringsbeginsel bleek dat dit met name raakt aan de uitlegbaarheid van de door het ML-algoritme gezette stappen om tot een bepaalde uitkomst te komen. Zoals hierboven toegelicht, is het ebm-algoritme beter uitlegbaar dan het xgb-algoritme. Beide algoritmes zijn echter tot op zekere hoogte alsnog een *black box*. Individuele selectiebeslissingen met behulp van een ebm- of xgb-algoritme kunnen door de gemeente alleen worden gemotiveerd aan de hand van statistische termen, die voor experts al ondoorzichtig kunnen zijn, laat staan voor de inwoner die de beslissing wil kunnen aanvechten. Het is daarmee maar zeer de vraag hoe uitlegbaar het ebm-algoritme daadwerkelijk is. De wens om de berekeningen van het algoritme in 'natuurlijke taal beschikbaar te stellen'[29] stuit hier op de grenzen van de statistische realiteit.

3.2.2. Zorgvuldigheid

Zoals we hebben geconstateerd vereist het zorgvuldigheidsbeginsel dat relevante feiten en belangen bekend zijn en met behulp van een geschikte methode worden afgewogen. Voor ML-toepassingen werpt dit vragen op over de volledigheid en correctheid van de gebruikte data en de geschiktheid van de gebruikte ML-methode. In de eerste plaats spelen risicobeheersingsmaatregelen hier een belangrijke rol, waarvoor onderhand verschillende richtlijnen zijn gepubliceerd.[30]

De wens om de berekeningen van het algoritme in 'natuurlijke taal beschikbaar te stellen' stuit hier op de grens van de statistische realiteit

Het vaststellen van een ML-toepassing als 'geschikte methode' is ondanks handreikingen en risicobeheersmaatregelen echter een lastig karwei. Op fundamentele rechten gerichte handvatten, zoals het IAMA en de 'Uitgangspunten voor (semi-)geautomatiseerde besluitvorming' van het College voor de Rechten van de Mens, zijn behoorlijk abstract en procedureel van aard.[31] Ze schrijven voor hoe organisaties de impact van algoritmes op de mensenrechten in kaart kunnen brengen om de evaluatieve afweging te kunnen maken. De normatieve beslechting van een dergelijke afweging wordt niet aangereikt binnen bovengenoemde *soft law*-kaders. Dit heeft een goede reden, omdat de juiste afweging hierin altijd contextafhankelijk is en niet in materiële zin door algemene kaders kan worden beslecht. Of en zo ja welke vorm van ML *in casu* de geschikte methode is om feiten en belangen af te wegen blijft daarmee een open vraag.

De casus Rotterdam leert dat er kwesties zijn waar relatief eenduidige oplossingen voor bestaan: de trainingsdataset bleek niet-representatief, waarmee de eis is geschonden dat feiten volledig bekend moeten zijn en correct worden afgewogen.[32] Maar de invulling van het zorgvuldigheidsbeginsel is een stuk minder duidelijk bij andere kwesties die uit de casus naar voren komen: zoals de vraag welke van de 60 beschikbare kenmerken geschikt zijn om als input te dienen voor een risicomodel. Is het aantal kinderen of de door een ambtenaar gemeten assertiviteit geschikt als selectiecriterium of niet? Moet men daarbij uitgaan van het statistische feit in hoeverre ze voorspellende waarde hebben? Of moet onafhankelijk van de effectiviteit van kenmerken een kwalitatieve afweging worden gemaakt welke intrinsiek (on)geschikt zijn? Evenmin eenduidig is de kwestie van de keuze voor een bepaalde ML-methode. In de casus Rotterdam is voor het xgb-algoritme gekozen uit een handvol alternatieven, omdat dit type algoritme als het meest effectief uit de testfase kwam.[33]

26. Documentatie over dit algoritme is te vinden in het Amsterdamse Algoritmeregister, zie: algoritmeregister.amsterdam.nl/ai-system/onderzoekswaardigheid-slimme-check-levensonderhoud/.

27. Zie eindevaluatie van de pilot met gedigitaliseerde fraudepreventie (TKN8), amsterdam.raadsinformatie.nl/vergadering/1203734/Raadscommissie%20Sociaal%2C%20Economische%20zaken%20en%20Democratisering%2014-02-2024.

28. De onder statistici populaire LIME en SHAP-values sorteren onvoldoende effect. Zie bijvoorbeeld D. Vale, A. El-Sharif & M. Ali, 'Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law', *AI Ethics* 2022, 2, p. 815-826.

29. Supra noot 17.

30. Zie bijvoorbeeld het Onderzoekskader Algoritmes van de Audit Dienst Rijk (2023) rijksoverheid.nl/documenten/rapporten/2023/07/11/onderzoekskader-algoritmes-adr-2023 en het Toetsingskader Algoritmes van de Algemene Rekenkamer (2021) rekenkamer.nl/onderwerpen/algoritmes-digitaal-toetsingskader.

31. Impact Assessment Mensenrechten en Algoritmes (IAMA), 31 juli 2021, rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes; College voor de Rechten van de Mens, 'Uitgangspunten voor (semi-)geautomatiseerde besluitvorming', 9 februari 2021, publicaties.mensenrechten.nl/publicatie/1980e51e-bb12-4bb1-8a9b-26c7a3aa2b86.

32. Zie lighthousereports.com/suspicion-machines-methodology/.

33. Wet openbaarheid van bestuur (Wob) verzoek van VPRO Argos/Lighthouse Reports, 2017020 Privacy Impact Assessment pilotfase Project Uitkeringsonrechtmatigheid, www.vpro.nl/argos/lees/onderwerpen/artikelen/2021/in-het-vizier-van-het-rotterdamse-algoritme-.html.

Verdere motivering van de keuze voor dit model ontbreekt.

Uit de casus Amsterdam blijkt een grotere zorgvuldigheid bij de ontwikkeling van het algoritme. In het publieke algoritmeregister staat dat voor het ebm-algoritme is gekozen vanwege het uitlegbare karakter.[34] Daarnaast is toegelicht welke variabelen wel en niet aan het algoritme zijn gevoed en waarom, en is een biasmeting uitgevoerd. Maar waarom er gekozen is voor ML-gedreven risicoprofilering, en waarom beter uitlegbare algoritmes (zoals regelgedreven algoritmes) niet zijn overwogen, blijft onduidelijk. Ook hier blijven dus vragen open die betrekking hebben op het zorgvuldigheidsbeginsel.

3.2.3. Fair play

Hierboven hebben wij opgemerkt dat uit het beginsel van *fair play*, als het in de context van besluitvorming waarin ML-gedreven risicoprofilering een rol speelt wordt geplaatst, mogelijk een inspanningsverplichting om algoritmische *bias* te voorkomen volgt.[35] Deze kwestie is relevant voor de casus Rotterdam, omdat is aangetoond dat de trainingsdata niet-representatief waren ten aanzien van jongeren, waardoor het model een vooringenomenheid kon ontwikkelen. Het beginsel van *fair play* (naast andere abbb en normen rond non-discriminatie die hier zeker ook van toepassing zijn) vereist dat zulke vooringenomenheden in datasets worden gemonitord en bestreden. Het wordt echter complexer als we ons richten op andere aspecten in de casus rondom *fairness* en vooringenomenheid.

Onafhankelijk van de kwaliteit van de dataset kan het model namelijk ook *bias* ontwikkelen door onderscheid te maken op kenmerken die ogenschijnlijk onschuldig zijn, maar sterk gecorreleerd zijn aan beschermde discriminatiegronden: het probleem van indirecte discriminatie door proxyvariabelen. In beide casussen doet deze kwestie zich voor met betrekking tot gebruikte kenmerken voor het algoritmische risicoprofiel. Het Rotterdamse algoritme gebruikte kenmerken zoals laaggeletterdheid of postcode die relatief sterk gecorreleerd zijn aan migratieachtergrond.[36] Maar vanuit statistisch oogpunt geldt dat alle mogelijke kenmerken in zekere mate zullen correleren aan beschermde gronden. Er bestaan hier kortom slechts graduele verschillen en er lijkt geen manier voorhanden om het probleem van proxyvariabelen op een zuivere manier te beslechten.

Boven op de kwestie van (proxy)discriminatie speelt bij het *fair play*-beginsel ook de bredere kwestie die we onder het zorgvuldigheidsbeginsel bespraken: welke vormen van onderscheid zijn acceptabel bij een ML-algoritme? Is het eerlijk om inwoners te profileren op basis van hun professionele voorkomen, ingevuld door een ambtenaar na een contactmoment? Zowel het subjectieve karakter van een dergelijke vaststelling als het mogelijk oneerlijke karakter om op basis van zulke persoonskenmerken onderscheid te maken, maakt een dergelijke praktijk twijfelachtig. Maar ook hiervoor is geen eenduidige norm beschikbaar.

Deze casusbespreking van gemeentelijke risicoprofilering laat enerzijds zien dat dit type ML-toepassingen het type problemen veroorzaakt en het soort vraagstukken oproept die de abbb zouden moeten oplossen of voorkomen. Anderzijds zien we dat de concrete invulling van de beginselen al snel op grenzen stuit. Met de opkomst van ML-algoritmen komt de schijnwerper te staan op de 'geschiktheid van de methode' in algemene zin. Tegelijkertijd is de algemene vraag naar de geschiktheid van ML als methode (waaronder we ook de uitlegbaarheid en eerlijkheid kunnen scharen) altijd context-afhankelijk en kan het niet goed door een algemeen kader worden voorgeschreven. Voor de vaststelling van de geschiktheid is dus aanvullende en casusgedreven normering nodig om werkelijk invulling te geven aan de abbb. Algoprudentie is naar ons oordeel een veelbelovend mechanisme om aan deze behoeften te voldoen.

4. Algoprudentie gedemonstreerd in de praktijk

Wat ontbreekt om verantwoorde inzet van algoritmische risicoprofilering te kunnen borgen zijn normen die tegelijkertijd flexibel en concreet zijn. Wij introduceren algoprudentie als manier om tot zulke flexibele en concrete normen te komen in het gebruik van ML-algoritmes. Voordat wij dit nieuwe begrip verder juridisch inbedden, willen we de potentie van algoprudentie allereerst aantonen aan de hand van de besproken casus uit Rotterdam. Het gaat hier met name om de volgende aspecten waar bestaande kaders geen pasklare antwoorden op bieden:

- de geschiktheid van ML als risicoprofileringsmethode in bijstandsheronderzoek, tegenover alternatieven zoals handmatige (expert-gedreven) profilering of aselecte steekproeven;
- transparantie- en uitlegbaarheidsvereisten voor ML-gedreven risicoprofilering;
- bepalen welke variabelen wel en niet wenselijk zijn om te worden gebruikt voor risicoprofilering, o.a. met het oog op proxydiscriminatie.

Stichting Algorithm Audit (waar de auteurs in verschillende rollen bij zijn betrokken) heeft onlangs een advies uitgebracht met betrekking tot deze normatieve vraagstukken rondom de casus Rotterdam.[37] Dit advies is uitdrukkelijk bedoeld als bijdrage aan algoprudentie. De bijdrage aan algoprudentie bestaat in totaal uit een probleemstelling, waarin de normatieve vraagstukken staan beschreven tegen de achtergrond van de relevante institutionele, juridische, en technische context, en een adviesdocument dat is opgesteld naar aanleiding van een deliberatief oordeel van een onafhankelijke commissie van experts en belanghebbenden. Wij beschrijven hier slechts kort en ter illustratie hoe deze algoprudentie een bijdrage kan leveren aan het beslechten van bovenstaande vraagstukken. Allereerst beantwoordt het advies (gedeeltelijk) bovenstaande vragen:

- Het stelt dat algoritmische risicoprofilering onder bepaalde strikte voorwaarden verantwoord kan worden ingezet in de context van bijstandsheronderzoek. Parallel gebruik van meerdere selectiemethoden (zowel algoritmische en handmatige profilering als aselecte steekproeven) wordt wenselijk geacht.
- Met het oog op uitlegbaarheidsvereisten oordeelt het dat het door de gemeente Rotterdam gebruikte xgb-algoritme ongeschikt is als methode. Het stelt een concrete norm voor wat als een voldoende uitlegbaar algoritme geldt.

Voorbeeld van algoprudentie: overzicht van variabelen waarvan een adviescommissie heeft beoordeeld of ze geschikt zijn als selectiecriteria voor (ML-gedreven) risicoprofilering in bijstandsheronderzoek.

– Ter handreiking voor het bepalen van de geschiktheid van profileringscriteria geeft het een lijst van verantwoorde en onverantwoorde variabelen met bijbehorende grond (zie figuur 1).

Deze concrete normen zijn in het advies ingebed in een evaluatie van de institutionele en sociaal-maatschappelijke context van het ML-algoritme en van bijstandsheronderzoek. Het oordeel en de daaruit voortvloeiende normen zijn casus-specifiek en context-afhankelijk, waarmee de normatieve flexibiliteit is verankerd. Tegelijkertijd is het oordeel generaliseerbaar naar hieraan analoge contexten, waar bijvoorbeeld de hierboven besproken casus van de 'Slimme check levensonderhoud' van de gemeente Amsterdam onder zou vallen. De algoprudentie die is gecreëerd kan daarmee productief bijdragen aan verantwoord algoritmegebruik. Als een andere gemeente nadenkt over de inzet van ML-gedreven risicoprofilering, kan ze lering trekken uit het algoprudentiële oordeel over de casus Rotterdam, om dat vervolgens mutatis mutandis op de eigen context toe te passen; iets waarvan wij op anekdotische basis kunnen vaststellen dat dit al gebeurt. Niet elk bestuursorgaan hoeft zo opnieuw het wiel uit te vinden, maar kan zich oriënteren op een bestaand, gemotiveerd oordeel. Op deze manier ontstaan gedeelde maar toch flexibele normen die het algoritmegebruik in een bepaalde context harmoniseren; *in casu* ML-risicoprofilering in de context van gemeentelijke voorzieningen.

5. Algoprudentie juridisch ingebed

In het voorgaande hebben we de werking en meerwaarde van algoprudentie trachten aan te tonen met een praktijkvoorbeeld. We willen tot besluit dit nieuwe begrip verder uitwerken en toelichten hoe dit als aanvullend instrument kan worden ingebed in het juridische landschap.

5.1. Algoprudentie als praktijk

Wij introduceren het begrip 'algoprudentie' voor de praktijk van concrete op casus gebaseerde en gedecentraliseerde oordeelsvorming over verantwoorde inzet van algoritmes. Hoe algoprudentie functioneert als praxis en hoe het algoprudentiële corpus exact tot stand komt zijn te grote vragen om hier volledig te beantwoorden. Het antwoord moet zeker ook niet alleen van de eerste exercities op dit vlak van Stichting Algorithm Audit afhangen. In de basis dient het een voortgaand gesprek tussen diverse partijen in de samenleving te zijn, gebaseerd op casuïstiek, over de concrete beslechting van normatieve vraagstukken die zich voordoen bij de inzet van algoritmische toepassingen. Algoprudentie hoeft zich niet te beperken tot het bestuursrechtelijke domein, maar kan in principe gelden voor alle (private en publieke) domeinen waar normatieve vraagstukken opkomen rondom de toepassing van algoritmes die niet door technisch-pragmatische oplossingen of juridische kaders worden beantwoord. De basis van algoprudentie wordt gevormd door de transparante publicatie van op casus gebaseerde uitspraken, die bestaan uit een toelichting van het normatieve vraagstuk inclusief context, en een gemotiveerd oordeel. In theorie kan algoprudentie net als jurisprudentie aanleiding geven tot annotaties van oordelen, die op een transparante en deliberatieve manier de collectieve oordeelsvorming verder ontwikkelen.

Hoe deze decentrale oordeelsvorming wordt vormgegeven en wie mag optreden als oordelende instantie zijn vooralsnog open vragen. Volledige decentralisatie is een mogelijkheid, waarbij het iedereen is toegestaan een bij-

34. Supra noot 26.

35. Bij het bepalen van bias in het algoritme-gedreven selectieproces dient ook het alternatief, bijvoorbeeld bias in een handmatig selectieproces, te worden onderzocht. Zie parool.nl/columns-opinie/opinie-onderzoek-vooringenomenheid-van-zowel-algoritme-als-ambtenaar~bd69aa5e/.

36. Zie ook sectie 4-5-4 van *Gekleurde technologie*, Rotterdamse Rekenkamer 2021, rekenkamer.rotterdam.nl/onderzoeken/algoritmes/.

37. Stichting Algorithm Audit, 'Risicoprofilering heronderzoek bijstandsuitkering' (AA:2023:02), 2023, algorithmaudit.eu/nl/algoprudence/#risk-profiling-social-welfare.

drage te leveren, ook al zullen uitspraken van bepaalde instanties meer gewicht hebben dan anderen. Beperkte decentralisatie is een andere mogelijkheid, waar het vooraf aangemelde en erkende instanties zijn die uitspraken kunnen doen (zoals ethische adviescommissies naast toezichthouders en andere formele organen). Afhankelijk van de vorm van decentralisatie zal het coördinerende instanties vereisen, en een bepaalde mate van standaardisering van wat als algoprudentie geldt en hoe dit gepubliceerd wordt. Algoprudentie kan op termijn het voorbeeld volgen van de ECLI-nummering en een internationaal gestandaardiseerde codering aannemen.[38]

Aan welke randvoorwaarden bijdragen aan algoprudentie moeten voldoen is evenzeer een open vraag. Oordelen hoeven niet per se tot stand te komen op de wijze die Stichting Algorithm Audit voorstaat, namelijk een deliberatief oordeel afkomstig van een commissie met inbreng van academische experts, belanghebbenden en betroffen groepen, al biedt deze specifieke aanpak wel voordelen. Ongeacht de wijze van institutionalisering zal het draagvlak voor de oordeelsvorming wel altijd afhangen van de zorgvuldigheid en geschiktheid van de gevolgde procedure.

5.2. Algoprudentie tegenover alternatieve instrumenten

Een in het oog springende kwestie bij algoprudentie is de status en legitimiteit van de daaruit voortvloeiende normen. Het moge duidelijk zijn dat algoprudentie als decentraal proces van normvinding geen direct bindend karakter kan hebben. Een logisch bezwaar is dan of nadere normen voor de inzet van ML-algoritmes niet beter in bindende wet- en regelgeving vastgelegd kunnen worden. Hoewel de regelgeving zeker concreter kan, zou codificatie van veel van de algoprudentiële normen niet passend zijn. Om een voorbeeld te geven: een wettelijk verbod op het gebruik van xgb-algoritmes zou niet goed passen in het bestuursrechtelijk stelsel, noch in de risicogestuurde aanpak van de AI-verordening, en te rigide zijn. Er zijn immers situaties denkbaar waarin een afweging toch ten gunste van deze techniek uitvalt, of waarin verdere technische ontwikkeling de problemen rond uitlegbaarheid ondervangt.

Ook al zijn algoprudentiële normen niet bindend, op verschillende manieren kunnen ze toch effect sorteren. Ten eerste zal er een zelfregulerend effect optreden als professionals op de hoogte zijn van de consensus in de algoprudentie op een bepaald gebied. Ze zullen dan goede redenen moeten hebben indien ze van de norm afwijken. Ten tweede kan dit effect versterkt worden door kaderende wetgeving, die bijvoorbeeld door documentatieverplichtingen kan afdwingen dat een organisatie motiveert welke afweging is gemaakt met betrekking tot een datagerelateerd systeem, waarbij de *state-of-the-art* in beschouwing moet worden genomen. Ten derde kunnen algoprudentiële normen dienen als input voor positiefrechtelijke interpretatie van de abbb en andere wettelijke kaders. Als algoprudentie op een bepaald vraagstuk voorhanden is, zal een rechter eerder geneigd zijn juridische consequenties te verbinden aan het niet-benutten hiervan, dan wanneer een bestuursorgaan in het duister moet tasten. We onderscheiden nog een vierde manier waarop algoprudentie effect sorteert: via politieke besluitvorming. Algoprudentie kan worden ingezet in het politieke speelveld om het functioneren van een bestuursorgaan kritisch te bevragen en te evalueren aan de hand van een onafhankelijke standaard.[39]

Concrete normen voor ML-algoritmes zullen ook voortvloeien uit de Europese AI-verordening. De AI-verordening is geënt op het reguleringsmodel van productveiligheid. De conformiteitsbeoordeling van hoog-risico AI-systemen zal daarom net als bij andere EU-productwetgeving gebeuren aan de hand van geharmoniseerde normen (zoals CEN- en ISO-normen). De AI-verordening zal de noodzaak van algoprudentie echter niet wegnemen. Ten eerste volgt uit het gegeven dat de AI-verordening over 'productveiligheid' gaat dat de geharmoniseerde normen een overwegend technisch karakter zullen hebben. Voor zover ze raken aan fundamentele rechten zullen ze procedureel van aard zijn.

Ten tweede, zoals we in de casussen hierboven hebben aangetoond, zijn de concrete evaluatieve afwegingen die moeten worden gemaakt sterk context-afhankelijk en kunnen ze dus niet met een generieke geharmoniseerde norm worden beslecht. Ten derde zijn geharmoniseerde normen niet publiek: ze zijn door veelal private organisaties opgesteld en alleen tegen betaling verkrijgbaar.[40] Geharmoniseerde normen kunnen dus in tegenstelling tot algoprudentie nooit de rol van publieke kennisopbouw en transparante collectieve oordeelsvorming vervullen.

6. Conclusie

We betogen dat bestaande juridische kaders, zoals de abbb, specifieke vraagstukken rondom de inzet van ML-algoritmes niet concreet genoeg normeren. Zoals onze bespreking van de Rotterdamse en Amsterdamse casus laat zien, bestaat er behoefte aan nadere invulling om aan onder andere de beginselen van motivering, zorgvuldigheid en *fair play* te voldoen. Met de introductie van het begrip algoprudentie stellen wij een aanvullend instrument voor om de normatieve lacune voor algoritmische toepassingen te vullen door middel van op casus gebaseerde, decentrale oordeelsvorming.

Algoprudentie houdt de belofte in dat organisaties die gebruik maken van algoritmische toepassingen worden voorzien van concrete normen hoe om te gaan met specifieke vraagstukken waar geen technisch-pragmatische, noch een eenduidige juridische oplossing voor bestaat. In dit artikel betogen wij dat er een eigen plek bestaat voor dit type algoprudentiële normen naast andere instrumenten en bronnen. De ontwikkeling van algoprudentie moet uiteraard nog tot wasdom komen. Door algoprudentie als begrip te introduceren hopen we echter zowel een impuls te geven aan de daadwerkelijke ontwikkeling van algoprudentie, als een eerste aanzet te doen voor de juridische inbedding ervan. •

38. Stichting Algorithm Audit gebruikt bijvoorbeeld de codering 2023:AA:02 waar de termen respectievelijk betrekking hebben op het jaar, de organisatie en het casusnummer, en de probleemstelling (2023:AA:02:P) of het adviesdocument daarvan (2023:AA:02:A). Met dank aan Martijn Staal voor het idee.

39. Zie bijvoorbeeld vragen uit de Amsterdamse gemeenteraad over het ebm-algoritme, amsterdam.raadsinformatie.nl/document/13573898/1/236+sv+Aslami%2C+IJmker+en+Garmy+inzake+toegepaste+profileringscriteria+gemeentelijke+algoritmes.

40. Al zou een recent arrest van het Hof van Justitie EU hier verandering in kunnen brengen: HvJ EU 5 maart 2024, ECLI:EU:C:2024:201.