Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses.

Inleiding

Doel van de richtlijnen is het geven van handvatten ten behoeve van *het ontwikkelen en het gebruiken van algoritmen door de overheid* en *ten behoeve van de publieksvoorlichting daarbij door overheden*.

Datum
1 maart 2021

Ons kenmerk
Richtlijnen Algoritmen

Aard circulaire
Informatie

Geldig van/tot

De richtlijnen komen voort uit de beschikbaarheid van nieuwe technieken om data-analyses uit te voeren. Overheden gebruiken al langer algoritmen waarvan de uitkomst wordt gebruikt als ondersteuning of advies voor verder onderzoek, beleid of uitvoeringstrajecten. Algoritmen zijn dus niet nieuw. Wél zijn er steeds meer gegevens digitaal beschikbaar voor overheden en zijn er steeds meer nieuwe technieken om data-analyses uit te voeren, onder andere gebruik makend van Artificial intelligence (AI).

Deze richtlijnen zijn een vervolg op de richtlijnen die met de brief van 8 oktober 2019 over waarborgen tegen risico's van data-analyses aan de Tweede Kamer zijn aangeboden.[1] Zoals in deze brief is aangekondigd, zijn de richtlijnen in verschillende trajecten op hun effectiviteit en uitvoerbaarheid getoetst en vervolgens geëvalueerd. Dit heeft geleid tot een aanpassing van de richtlijnen.

De richtlijnen zien toe op de ontwikkeling en operationele inzet van algoritmen. Het risico van inzet van het algoritme wordt met name bepaald door een samenspel van gebruikte gegevens de bedrijfsprocessen en systemen op de werkvloer. Het belangrijkste wat met deze richtlijnen wordt geïntroduceerd, is het creëren van transparantie en de waarborgen om mogelijke risico's van de inzet van algoritmen te mitigeren. Tevens introduceren de richtlijnen de vastlegging van de procesmatige totstandkoming en implementatie als concrete bewijsstukken van zorgvuldig handelen, zowel bij de ontwikkeling als bij de inzet van een algoritme.

Hiermee geven de richtlijnen een invulling van de algemene beginselen van behoorlijk bestuur bij nieuwe vormen van bedrijfsprocessen die door inzet van algoritmen en data-analyse ontstaan. De richtlijnen stellen de invloed van het algoritme in de uitvoering van een publieke taak centraal. Daarmee ontstaat tevens aandacht voor:

a. de operationele inzet van algoritmen t.b.v. data-analyses op data/gegevens waarvoor de grondslag voor gebruik daarvan niet aanvullend hoeft te worden getoetst;
b. de inzet van algoritmen die geen persoonsgegevens verwerken, maar toch de belangen van burger, bedrijf en maatschappij kunnen schaden.

De aandacht voor data-analyses op de werkvloer leidt tot een verhoogd risicobewustzijn binnen de (hiërarchische structuur van de) organisatie en tot meer behoefte aan samenwerking in multidisciplinaire teams en (intercollegiale) uitlegbaarheid zijnde één van de doelen van de richtlijn. De richtlijnen zijn in algemene zin van toepassing op algoritmen met impact op burger, bedrijf of samenleving. Het gaat daarbij dus niet alleen om algoritmen die persoonsgegevens verwerken, maar om data-analyses in de brede zin.

De richtlijnen zijn vooral gericht op de transparantie en daarmee de uitlegbaarheid van algoritmen, de werking en toepassing daarvan, bedoeld om het inzicht te vergroten alsmede de kwaliteit en betrouwbaarheid van algoritmen te verbeteren. Hiertoe bevatten de richtlijnen vereisten met betrekking tot:

1. Bewustzijn risico's
2. Transparantie & Uitlegbaarheid
3. Gegevensherkenning
4. Auditeerbaarheid

[1] Kamerstukken II 2018/19, 26643, nr. 641

5. Verantwoording
6. Validatie
7. Toetsbaarheid
8. Publieksvoorlichting

De vereisten 1 tot en met 7 zijn opgenomen in de 'Richtlijn voor het toepassen van algoritmen door overheden'; Deze zijn op basis van concrete aanwijzingen in Deel 2 van de richtlijnen opgenomen, waarbij de vereisten gekoppeld zijn met het normenkader waarop de Auditdienst Rijk de toepassing van AI door de Rijksoverheid auditeert.

Het vereiste onder 8 is het onderwerp van de 'Richtlijn inzake publieksvoorlichting over data-analyses' en betreft de voorlichting aan het publiek over data-analyses en de informatieverschaffing door een overheidsdienst. Deze is eveneens in Deel 2 in meer detail uitgewerkt.

Leeswijzer
Dit document bestaat uit twee delen, waarbij het eerste deel de kaders voor de toepassing van de richtlijnen bevat en het tweede deel, de daadwerkelijke richtlijnen voor algoritmen en publieksvoorlichting. Het tweede deel richt zich primair op de functionarissen die bij de ontwikkeling en het beheer van algoritmen een rol vervullen respectievelijk de communicatiemedewerkers die verantwoordelijk zijn voor de publieke informatievoorziening. Het eerste deel is, naast voornoemde functionarissen, ook van belang voor de andere functionarissen die, vanuit hun rol of discipline, eveneens betrokken zijn bij de ontwikkeling en toepassing van algoritmen (zie hierover paragrafen 3 en 5), en is als volgt opgebouwd. Als eerste wordt een definitie gegeven van algoritmen, en worden de verschillende typen en wijzen waarop algoritmen kunnen worden ingezet beschreven (paragraaf 1); vervolgens wordt ingegaan op het belang van transparantie en het verschil tussen technische transparantie en uitlegbaarheid (paragraaf 2). In de paragrafen 3 tot en met 8 wordt aandacht besteed aan de positionering en het karakter van de richtlijnen, hoe die zich verhouden tot andere instrumenten, voor wie ze bedoeld zijn, wanneer de richtlijnen van toepassing zijn en hoe ze gemonitord worden.

Deel I - kaders voor de toepassing van de richtlijnen

1. Definitie, typen en inzet van algoritmen

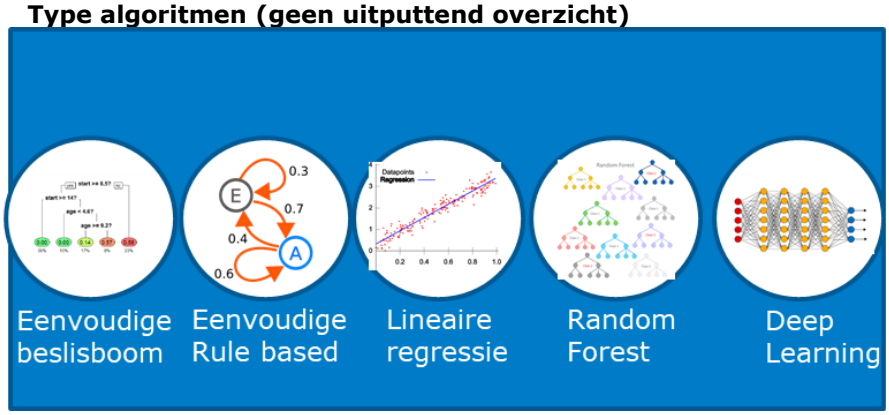
Definitie
Er bestaan verschillende definities van wat onder een algoritme wordt verstaan.
In dit document wordt de volgende definitie van een algoritme gehanteerd.

Algoritme: Een wiskundige formule of model, dat wordt uitgevoerd middels een computer, om een probleem op te lossen, een vraag te beantwoorden, een voorspelling te doen, een beslissing te nemen of die te ondersteunen.

Aan de hand van data die door het algoritme worden gebruikt, wordt in verschillende stappen toegewerkt naar het beoogde eindresultaat, bijvoorbeeld het toekennen van een vergunning. Op basis van dat resultaat kan men dan actie ondernemen (al dan niet toekennen van de vergunning). De precieze stappen die worden doorlopen verschillen per algoritme en zijn afhankelijk van de wijze waarop het algoritme wordt ingezet.

Typen algoritmen
Er bestaan verschillende typen algoritmen. Deze kunnen variëren van een beslisboom met een beperkt aantal variabelen tot complexe en zelflerende algoritmen, zoals zogenaamde '*machine learning'* of '*deep learning'* algoritmen. Deze laatste twee varianten worden o.a. ook gebruikt bij toepassingen van Artificial Intelligence (AI).[2] Hiermee kunnen complexe verbanden in data worden vastgesteld die een mens zelf moeilijk of niet kan vinden. Afhankelijk van de toepassing kan het redeneerproces van de AI soms inherent ondoorzichtig zijn en voor een mens lastig of niet te begrijpen.

Type algoritmen (geen uitputtend overzicht)



Inzet van algoritmen
Naast de definitie en de typen algoritmen is ook van belang hoe en waarvoor het algoritme wordt ingezet (welk doel). We kunnen onderscheid maken in de volgende vier "inzetgebieden":

1. Beschrijvend – Analyse van 'Wat gebeurt er'– het algoritme geeft een weergave van wat er wordt waargenomen;
2. Diagnostisch – Analyse van 'Waarom gebeurt het' – het algoritme geeft een waarschijnlijkheid of analyse van waarom iets optreedt, veroorzaakt door positieve, negatieve of predictieve waarden, of berekent een bepaalde waarschijnlijkheid die wordt gebruikt in werkprocessen en/of besluitvorming;

[2] In het Strategisch Actieplan AI (SAPAI) wordt KI als volgt gedefinieerd: 'Er is geen algemeen geldige definitie van AI die consistent wordt gebruikt door alle belanghebbenden. Wij gebruiken de omschrijving van AI door de Europese Commissie: "AI verwijst naar systemen die intelligent gedrag vertonen door hun omgeving te analyseren en - met een zekere mate van zelfstandigheid - actie ondernemen om specifieke doelen te bereiken.' Zie Kamerstukken II 2018/19, 26643, nr. 640, p. 9.

3. Voorspellend – Analyse van 'Wat kan er gebeuren' - het algoritme geeft een verwachting af, voorspelling van wat kan gebeuren of een kans/waarschijnlijkheid van een toekomstige handeling of gebeurtenis;
4. Voorschrijvend – Analyse van 'Wat moet er gebeuren' - Het algoritme bepaalt en/of dicteert de beslissing/actie of uitvoering.

De inzet van het algoritme is relevant voor de impact op de burger of de impact die het heeft op de te ondersteunen beslissing: een voorschrijvend algoritme heeft in de regel meer impact dan een beschrijvend algoritme (wanneer toegepast op dezelfde casus).[3] Bij een voorschrijvend algoritme is veelal sprake van besluitvorming, hetgeen niet het geval is bij beschrijvend, diagnostisch, of voorspellend algoritme. Zie over impact, onder paragraaf 7.

Naast de impact neemt ook de autonomie van het algoritme in relatie tot het inzetgebied van het algoritme toe (ook hier geredeneerd uit optiek van dezelfde casus). Wanneer een data-analyse leidt tot een voorschrijvende uitkomst (hoge autonomie), bijvoorbeeld bij geautomatiseerde besluitvorming, zal in die analyse in een aantal gevallen ook een beschrijvende, diagnostische en eventueel een voorspellende analyse plaatsvinden. Dat is immers nodig om adequaat te analyseren wat de beste maatregel of beslissing is.

Als voorbeeld: indien automatisch (voorschrijvend) een 'boete' wordt toebedeeld, zal eerst bepaald moeten worden wat er gebeurt (vaststelling van de overschrijding van de snelheidslimiet met x km per uur), analyse van 'waarom' het gebeurt (binnen of buiten de bebouwde kom) met welk voertuig (motor, vrachtwagen, auto, auto met caravan, met aanhanger, met fietsendrager) alvorens de 'boete' automatisch kan worden 'uitgeschreven'. Hierbij zou de 'boete' ook nog afhankelijk kunnen zijn van analyse van 'wat kan er gebeuren', bijvoorbeeld door de kans van herhaling van overtreding door bestuurder op basis van zijn/haar historisch gedrag mee te wegen in het besluit.

[3] Afhankelijk van de toepassing en het doel, kan het ook zo zijn dat een beschrijvend algoritme meer impact heeft dan een voorschrijvend algoritme.

2. Transparantie; uitlegbaarheid en technische transparantie

Belang van transparantie en uitlegbaarheid
Een belangrijk doel van de richtlijnen is het vergroten van de transparantie (zowel uitlegbaarheid als technische transparantie/werking) van algoritmen, de werking en de toepassing daarvan. Transparantie rond algoritmische data-analyses kan bijdragen aan het vertrouwen dat burgers in deze analyses hebben. In zoverre is transparantie vooral een middel, waar vertrouwen het doel moet zijn. Transparantie zal de burger, de interne en externe controleur, de toezichthouders en de rechter beter in staat stellen de werkwijze van de overheid bij een data-analyse te begrijpen en te toetsen. Daarmee draagt transparantie bij aan een zo evenwichtig mogelijke verhouding tussen burger en overheid. Transparantie kan ook leiden tot een betere naleving van wet- en regelgeving: het noemen van variabelen of drempelwaarden die de overheid bij data-analyses hanteert, kan enerzijds calculerend gedrag in de hand werken, maar kan er anderzijds juist ertoe leiden om bepaalde (ongewenste) acties van burger te voorkomen (functie van *nudging*).[4]

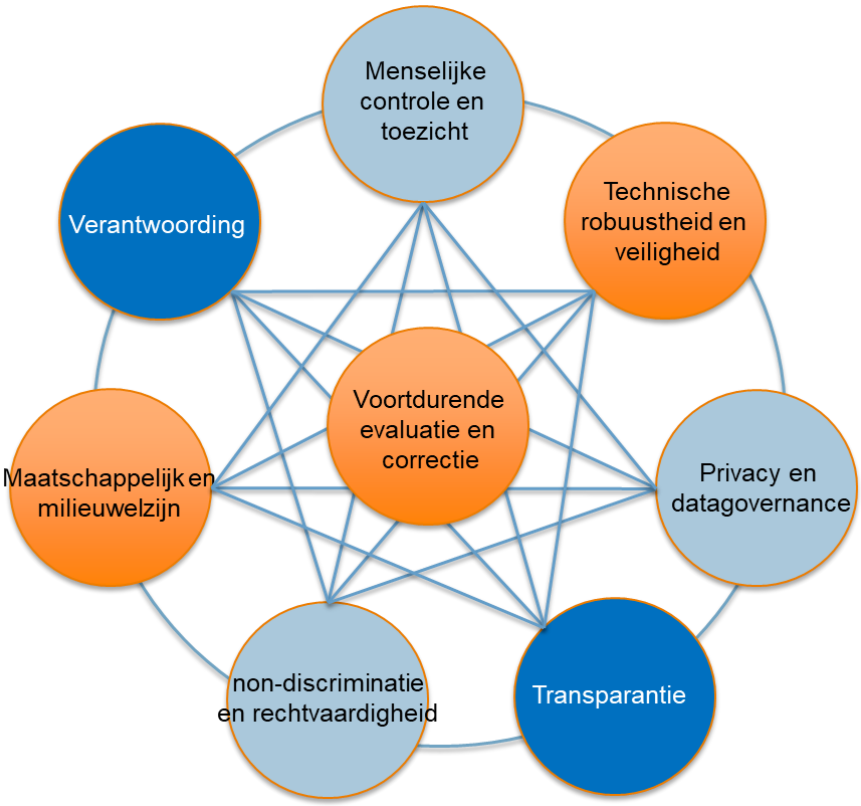
Waar de 'richtlijn voor het toepassen van algoritmen' zich richten op de transparantie binnen de organisatie en ten behoeve van de interne en externe controleurs, toezichthouders, rechter en de individuele burger[5], ziet de 'richtlijn inzake publieksvoorlichting' vooral op de transparantie naar buiten toe, richting het publiek.

Transparantie en uitlegbaarheid staat in nauw verband met het beginsel van verantwoording. Verantwoording is enkel mogelijk wanneer het algoritme voldoende transparant en uitlegbaar is omdat het alleen dan kan bijdragen aan de motivering van overheidshandelingen en -beslissingen. Voor die transparantie is het op haar beurt noodzakelijk dat het algoritme uitlegbaar en auditeerbaar is (proces). Ten behoeve van deze verantwoording via audits worden andersom eisen gesteld aan het algoritme op het punt van gegevensherkenning, validatie en toetsbaarheid (van het algoritme). Daarmee dient bij de inkoop van systemen dan wel de ontwikkeling van algoritmen al rekening te worden gehouden. Algoritmen dienen zogezegd *auditable by design* te zijn.

Transparantie en uitlegbaarheid kan niet los worden gezien van andere belangrijke vereisten voor algoritmen. Deze worden goed weergegeven in de Ethische Richtsnoeren voor betrouwbare AI opgesteld door de High Level Expert Group on AI, opgericht door de Europese Commissie[6]. Hoewel AI doorgaans gebruik maakt van specifieke algoritmen (zie hierboven typen algoritmen) en de richtlijnen daartoe niet beperkt zijn, geven de Ethische Richtsnoeren een helder overzicht van de belangrijkste aandachtspunten die voor het ontwikkelen van algoritmen van belang zijn, namelijk:

1) menselijke controle en menselijk toezicht;
2) technische robuustheid en veiligheid;
3) privacy en datagovernance;
4) **transparantie**;
5) diversiteit, non-discriminatie en rechtvaardigheid;
6) milieu- en maatschappelijk welzijn;
7) **verantwoordingsplicht**.

[4] Zie over het belang van transparantie van algoritmen, de werking en toepassing daarvan, de uitspraak in de zaak SyRi waarin de rechter oordeelt dat bij gebrek aan controleerbaar inzicht in de werking van het risicomodel, bijv. het type algoritme dat wordt gebruikt, de validatie van het risicomodel, de methode van risicoanalyse, de risico-indicatoren en de verificatie daarvan, of nog de objectieve feitelijke gegevens welke gerechtvaardigd tot de conclusie kunnen leiden dat sprake is van een verhoogd risico, de Syri-wetgeving over de inzet van Syri in strijd is met artikel 8, tweede lid EVRM (inmenging in het privéleven). Aldus valt volgens de rechtbank 'moeilijk in te zien hoe een betrokkene zich kan verweren tegen het gegeven dat ten aanzien van hem of haar een risicomelding is gedaan. […] Het recht op respect voor het privéleven houdt ook in dat een betrokkene in redelijke mate in staat moet worden gesteld zijn of haar gegevens te volgen. Het belang van transparantie, met het oog op controleerbaarheid, is mede zwaarwegend omdat aan het gebruik van het risicomodel en de analyse die in dat verband wordt verricht het risico verbonden is dat (onbedoeld) discriminerende effecten optreden'. Rb Den Haag, 5 feb. 2020, ECLI:NL: RBDHA:2020:865, r.o. 6.89 ev
[5] Naar analogie met de AVG (artikel 4, onder 1) wordt onder betrokkene verstaan: geïdentificeerde of identificeerbare personen wiens persoonsgegevens door algoritmen worden verwerkt.
[6] https://ec.europa.eu/futurium/sites/futurium/files/b1_download_guidelines.jpg. Zie hoofdstuk II.



De onderhavige richtlijnen leggen de focus primair op vereisten die verband houden met transparantie en verantwoording en aandacht aan de vereisten en waarborgen inzake non-discriminatie/diversiteit, privacy en menselijke controle en menselijk toezicht.

Algoritmen en transparantie

Gelet op hun technische aard zijn algoritmen niet altijd en niet voor iedereen transparant. Naarmate algoritmen complexer worden - waarbij zoals gezegd relevant is hoe algoritmen worden ingezet – kunnen deze ook minder doorzichtig zijn, al zijn er ontwikkelingen die dat inzicht proberen te verbeteren.

Algoritmen zullen wel ten alle tijden uitlegbaar moeten zijn.

Ondoorzichtigheid van algoritmen kan voortkomen uit de bescherming van eigendomsbelangen van hun makers. In andere gevallen wordt transparantie van algoritmen bewust achterwege gelaten, omdat getracht wordt 'gaming the system' te voorkomen en zo hun werkzaamheid te behouden.[7]

'gaming the system' is het verschijnsel dat burgers misbruik maken van de gegeven informatie en calculerend gedrag gaan vertonen, waardoor de effectiviteit van het overheidshandelen nadelig wordt beïnvloed. Daarom zal de informatieverschaffing achterwege of beperkt moeten blijven, voor zover een algemeen belang als bedoeld in artikel 23, eerste lid, AVG zich daartegen verzet[8] en rekening te houden met de specifieke bepalingen op genomen in artikel 23, tweede lid, AVG. Het gaat daarbij om belangen als de nationale of openbare veiligheid, economische en financiële belangen met inbegrip van fiscale aangelegenheden, volksgezondheid en sociale zekerheid, de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten en de taken op het gebied van toezicht en inspectie op genoemde terreinen.[9] De overheid moet die uitzonderingen goed onderbouwen.

[7] Gaming the system: wanneer (veel) inzicht in algoritmen en de werking daarvan wordt gegeven kunnen kwaadwillenden hier misbruik van maken.
[8] Zie ook artikel 21, tweede lid, Wjsg en 27, eerste lid, Wpg.
[9] Als het gaat om het belang van de nationale veiligheid, laat het Cybersecurity Beeld Nederland 2019 zien dat er een significante dreiging uitgaat van cybercriminelen en statelijke actoren. Zie Cybersecuritybeeld Nederland 2019, Kamerstuk 26 643, nr. 614. Hierom dienen organisaties in te zetten op het versterken van hun digitale weerbaarheid. Door het nemen van gepaste beheersmaatregelen, waar het kabinet op inzet middels de Nederlandse Cybersecurity Agenda, kan dit risico aanzienlijk worden verkleind. Zie hierover de Nederlandse Cybersecurity Agenda (Kamerstuk 26 643, nr. 536), Voortgangsrapportage NCSA (Kamerstuk 26 643, nr. 614). Dit kan uiteraard betekenen dat transparantie achterwege moet blijven.

Technische transparantie en uitlegbaarheid
Transparantie en uitlegbaarheid van dataverwerkingsprocessen is nodig voor effectieve controle en toezicht op de totstandkoming van data-analyses van de overheid en de uitkomsten daarvan. Het is van belang hierbij onderscheid te maken tussen "technische transparantie" en "uitlegbaarheid".

Onder **technische transparantie** verstaan we inzicht in de algoritmische methode die wordt toegepast (beslisboom, neuraal netwerk), de broncode, hoe het algoritme is getraind, als ook de gebruikte data, invoervariabelen, parameters en drempelwaarden die worden gebruikt etc.

Bij **uitlegbaarheid** gaat het om het in begrijpelijke taal kunnen uitleggen van de uitkomsten van data-analyses en hoe deze tot stand zijn gekomen en/of interpreteerbaar is. Vergelijkbaar met de uitleg die we krijgen van een mens die dezelfde beslissing neemt. Het betreft dan zowel intercollegiale uitlegbaarheid als uitlegbaarheid naar betrokkenen.

Technische transparantie resulteert niet altijd in uitlegbaarheid. Het redeneerproces van een algoritme is niet altijd even inzichtelijk en te doorgronden. Bij complexe algoritmen, zoals zelflerende, deep learning algoritmen, met een grote hoeveelheid variabelen en neurale lagen kan het zelfs voor experts lastig zijn om op basis van technische transparantie, het algoritme en de werking daarvan voldoende te doorgronden. Voor inherent ondoorzichtige modellen zijn wel technieken ontwikkeld, of in ontwikkeling, om achteraf te achterhalen op welke informatie een algoritme zijn uitkomst baseert. Tegen deze achtergrond wordt bij uitlegbaarheid van algoritmen de focus gelegd op het beschrijven van het doel dat met het algoritme wordt nagestreefd, de procedures die door het algoritme worden gevolgd, welke variabelen of beoordelingscriteria[10] doorslaggevend zijn geweest voor de uitkomst en het type gegevens dat wordt gebruikt (de kwaliteit en herkomst ervan, hoe de gegevens worden gecombineerd).[11]

Daarmee is uitlegbaarheid in veel gevallen veelzeggender dan technische transparantie.
Op haar beurt is (de mate van) uitlegbaarheid ook weer afhankelijk van het type algoritme, de wijze waarop het wordt ingezet en de betrokken functionaris(en):
Type algoritme: in het algemeen geldt dat hoe complexer het algoritme, hoe lastiger het is om de logica erachter begrijpelijk uit te leggen.
Inzet algoritme: voor een algoritme dat beschrijvend wordt ingezet, bijvoorbeeld voor het categoriseren van foto's, is een andere mate van uitlegbaarheid nodig dan voor een algoritme dat wordt ingezet voor de voorspelling van fraude of waar deze als uitkomst een besluit oplevert.

Een bijzondere situatie betreft algoritmen die ingezet worden ten behoeve van opsporing of vervolging en ten aanzien waarvan niet wenselijk is dat hun werking uitgelegd wordt; dit om ontwijkend gedrag te voorkomen en hun werkzaamheid te behouden. Uiteraard moeten deze algoritmen uitlegbaar zijn voor controllers, toezichthouders of rechters, en moet kunnen worden verantwoord waarom de geheimhouding noodzakelijk is en dient de burger inzicht te hebben onder welke omstandigheden en om welke redenen algoritmen worden ingezet.

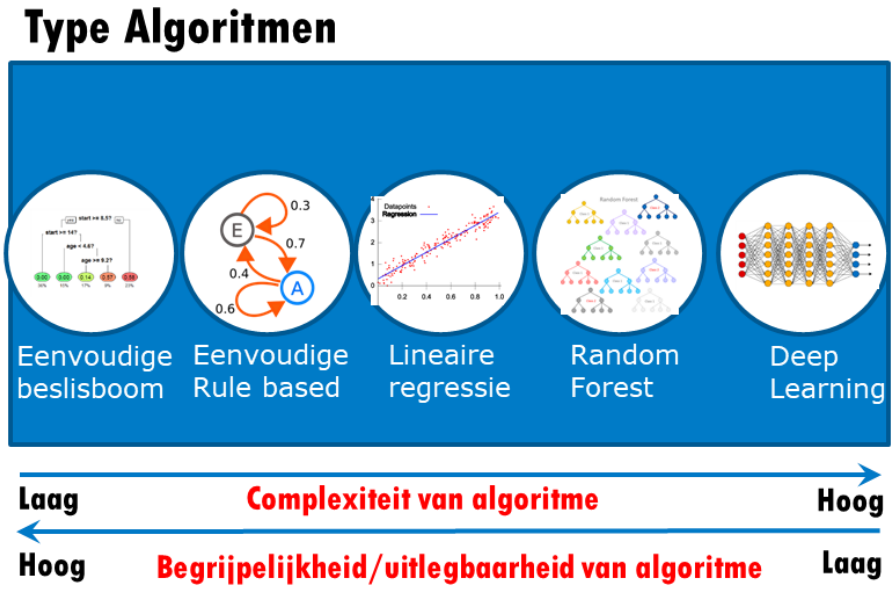
Meer algemeen geldt dat de overheid geen algoritmen hanteert waarvan de uitkomsten niet navolgbaar en controleerbaar zijn.

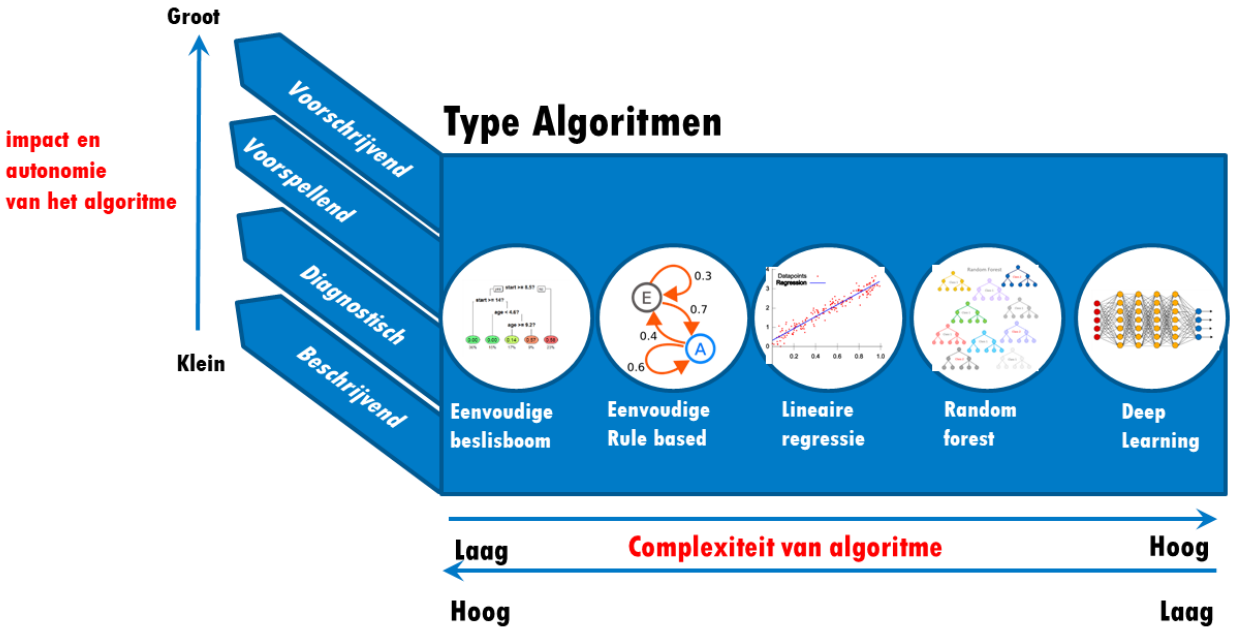
Functionaris: verschillende typen functionarissen hebben behoefte aan verschillende informatie; technische experts zullen vooral behoefte hebben aan informatie over de werking van het algoritme, terwijl (privacy)juristen of beleidsmedewerkers eerder willen weten in

[10] Bij transparantie en uitlegbaarheid zijn de *qualifiers* (variabelen en drempelwaarden) binnen een algoritme van groot belang. Welke *qualifiers* zorgen ervoor dat men tot een risicoprofiel komt, en kunnen deze *qualifiers* inzichtelijk worden gemaakt? Een toetsingscommissie zou aan de hand van *case studies* kunnen toetsen wanneer diensten over deze *qualifiers* wel en niet transparant kunnen worden gemaakt voor eventuele betrokkenen. Deze vorm van transparantie zou dan om *gaming the system* te voorkomen bij voorkeur niet vooraf in het proces moeten plaatsvinden maar achteraf. Zie ook voetnoot 2.
[11] Zie ook Kamerstukken II 2018/19, 26643, nr. 641 en Kamerstukken II 2018/19, 26643, nr. 570, blz. 4.
Zie over openbaarmaking van beslisregels bij algoritmen die een één-op-één-vertaling van wetgeving en beleidsregels vormen, de antwoorden d.d. 12 juni 2019 van de minister van Rechtsbescherming op de vragen van het lid Buitenweg (Groen Links) over de motivering van automatisch genomen besluiten. Aanhangsel Kamerstukken II, 2018/19, 3088.

welke wettelijke- en/of beleidscontext het algoritme wordt toegepast, en of het tot uitkomsten komt die zich verhouden met wetgeving en beleid.

In onderstaande figuren wordt de wisselwerking tussen de complexiteit van het algoritme en de mate van technische transparantie en uitlegbaarheid geïllustreerd. In het tweede figuur wordt dit verder gecombineerd met de wijze van inzet en autonomie van het algoritme en daarbij behorende impact. Over de impact, zie paragraaf 7. Uiteraard kunnen er ook zeer complexe beslisbomen bestaan, dus onderstaande is geredeneerd als vuistregel.





3. Positionering en karakter van de Richtlijnen

Positionering
Een algoritme fungeert niet zelfstandig maar wordt toegepast in het kader van een bepaalde context, het beleids- of uitvoeringsterrein van de organisatie die het algoritme toepast, en met een bepaald oogpunt gerelateerd aan de taak van de organisatie.[12] Het is onderdeel van een *socio-technisch systeem*.[13]

Het is dan ook belangrijk om bij de ontwikkeling en inzet van algoritmen, rekening te houden met de bredere bestuurlijke en organisatorische context waarin algoritmen worden toegepast. De ontwikkeling van algoritmen is met andere woorden geen autonoom of geïsoleerd proces maar een samenspel tussen verschillende disciplines (beleid, juridisch, ICT en beveiliging en bestuurlijk) waarbij iedere discipline een eigen belang en verantwoordelijkheid heeft. Zo dient het algoritme en de daarop gebaseerde data-analyse kenbaar en voorzienbaar zijn te voor burgers, dient het nodig te zijn voor de taakuitoefening, moeten de daarbij verwerkte gegevens rechtmatig zijn verkregen en de bijbehorende gegevensverwerkingen rechtmatig en proportioneel. Hierbij moeten burgers via reguliere processen een klacht of bezwaar kunnen indienen tegen de uitkomsten van de data-analyse.

Ondanks het belang daarvan zijn deze richtlijnen niet gericht op de vereisten rond het proces en de organisatie-inrichting. In de regel bestaan deze processen en organisatierichting namelijk al vanuit andere reguliere processen, zoals de juridisch-control functie, de beveiligingsfunctie, de privacy- en grondrechtelijke functie, de communicatiefunctie of de reguliere managementlijn.

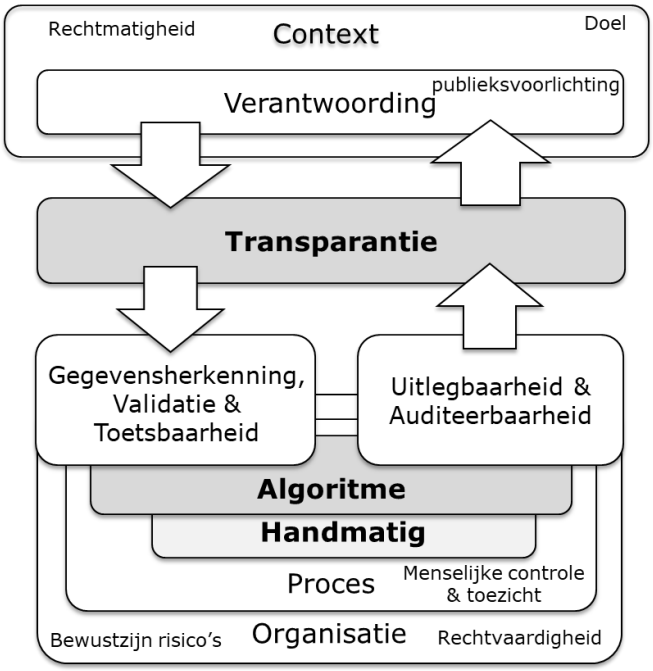
Voor geautomatiseerde processen op basis van algoritmen geldt geregeld dat deze processen in de plaats komen van handmatige processen waarvoor organisaties reeds een proces en organisatie-inrichting hebben, gelijk aan de andere processen of diensten die een organisatie levert.

Als voorbeeld: Indien de risicoclassificatie van een mogelijke fraudeur voorheen een handmatig proces was, gebaseerd op het onderbuikgevoel of gehanteerde kenmerken door een inspecteur/controleur, wat vervolgens na onderzoek tot inhouding van de uitkering leidt, dan zal een organisatie daar een proces en organisatie inrichting voor nodig hebben. Indien nu het handmatige proces (van risicoclassificatie) vervangen wordt door een geautomatiseerd proces middels inzet van een algoritme, zal duidelijk moeten zijn (transparant) hoe dat algoritme tot die beslissing of ondersteuning van die beslissing komt. Er is immers geen inspecteur/controleur die op dat moment bevraagd kan worden en/of uitleg kan geven. De opvolging van de geautomatiseerde risicoclassificatie zal daarentegen hetzelfde zijn als bij de handmatige selectie. Een behandeld ambtenaar zal er verdere opvolging aan moeten geven, er is immers geen sprake van een voorschrijvend/autonoom besluit. Hiervoor kan gebruik worden gemaakt van de reeds bestaande processen en organisatie-inrichting.

Het is aan organisaties om ervoor te zorgen dat de richtlijnen in de bestaande processen en structuren worden verankerd.

Onderstaande figuur geeft weer dat voor zowel een handmatig proces als het geautomatiseerde proces op basis van een algoritme dezelfde verantwoording en interne processen gelden.

[12] Zie hierover de definitie van Maranke Wieringa, die een algoritmisch systeem definieert als een socio-technische verzameling bestaande uit een combinatie van technische onderdelen, sociale praktijken en (organisatie)cultuur. Maranke Wieringa, 'What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability', in: Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, 2020, Barcelona, Spain. ACM, New York, NY, USA.
[13] https://dl.acm.org/doi/abs/10.1145/3351095.3372833



Karakter

Uit de aard van richtlijn vloeit algemeen voort dat zij geen dwingend karakter hebben. Wel is het principe '*comply or explain'* van toepassing op de richtlijnen. Dit principe houdt in dat als uitgangspunt geldt dat organisaties de richtlijnen moeten volgen, of nadrukkelijk uitleggen waarom zij ervan afwijken. Dit principe geldt in eerste instantie voor algoritmen die nieuw worden ontwikkeld, maar ook voor algoritmen die reeds toegepast worden. Wanneer van de richtlijnen wordt afgeweken wordt daarover binnen de hiërarchische structuur van de organisatie verantwoording afgelegd. Ten behoeve daarvan worden organisaties onder meer geacht de toepassing van de richtlijnen te verankeren in het kader van een P&C-cyclus, waarbij jaarlijks door organisaties wordt gerapporteerd over situaties waarin besloten is om af te wijken van de richtlijnen. Zoals in de Kabinetsreactie op het onderzoek 'Toezicht op het gebruik van algoritmen door de overheid' is toegezegd, wordt op dit moment verkend of, en onder welke voorwaarden, een vorm van rapportage door overheidsorganisaties over de door hen ingezette algoritmen wenselijk is (bv. middels algoritme register). E.e.a. zou betekenen dat organisaties jaarlijks zouden rapporteren over de door hen ingezette algoritmen en daarbij in het bijzonder over situaties waarin besloten is om af te wijken van de richtlijnen. Zie ook Deel II, Richtlijnen voor het toepassen van algoritmen door overheden, onder Verantwoording.

4. Hoe verhoudt de richtlijn zich tot andere instrumenten (wetgeving en beleid)

Richtlijnen voor het toepassen van algoritmen
De richtlijnen zijn bedoeld voor het geven van handvatten ten behoeve van het ontwikkelen en het gebruiken van algoritmen door de overheid. In die zin worden de richtlijnen gehanteerd naast, en zo mogelijk in afstemming met andere instrumenten en hulpmiddelen voor overheidsverwerkingen. De richtlijnen komen dus niet in de plaats van andere bestaande instrumenten. Het gebruik van algoritmen door de overheid dient, voor zover dat gebruik burgers raakt, steeds in overeenstemming te zijn met de grondrechten, hiervoor kan de Impact Assessment Mensenrechten en Algoritmen (IAMA) worden ingezet.

In het geval de toepassing van het algoritmische systeem bij wet wordt geregeld, zoals het PNR-systeem[14], of onderdeel uitmaakt van nieuw beleid kan worden gedacht aan instrumenten uit het IAK zoals de uitvoerbaarheids- en handhaafbaarheidstoets.

Gegevensbescherming
In het geval er gebruik wordt gemaakt van persoonsgegevens, dient uiteraard te worden voldaan aan de voorwaarden en beginselen uit de AVG of de Richtlijn gegevensbescherming opsporing en vervolging (hierna: de Richtlijn), waaronder het vereiste dat er voor de verwerking een rechtsgrond is.[15] Voor algoritmen die waarschijnlijk een hoog risico inhouden voor de rechten van betrokkenen zal ook een *Data Protection Impact Assessment* (hierna: DPIA) moet worden uitgevoerd (zie ook hieronder). Het vereiste inzake uitlegbaarheid dat beoogt de transparantie van het algoritmische proces te vergroten kan worden gezien als een nadere invulling - toegesneden op de specifieke kenmerken van algoritmische data-analyses - van het transparantiebeginsel uit de AVG.[16]
Waar relevant wordt in de richtlijnen verwezen naar artikelen uit de AVG en Richtlijn.

Beveiligingsrichtlijnen
Op het terrein van informatiebeveiliging kan verder worden gedacht aan de volgende normenkaders en instrumenten:

- het Voorschrift informatiebeveiliging Rijksdienst 2007 (VIR 2007);
- het Besluit voorschrift informatiebeveiliging Rijksdienst – bijzondere informatie 2013 (VIRBI 2013);
- de Baseline Informatiebeveiliging Overheid (BIO);
- de Business Impact Analyse (BIA).[17]

Zowel uit de BIO, VIR 2007 als uit de AVG en de Richtlijn kan worden afgeleid dat de verantwoordelijke, ter borging dat de beveiliging steeds adequaat is voor de huidige stand van de techniek en de organisatie een planning-en controlcyclus (plan-do-check-act) heeft ingericht.

Archiefwet
De informatie die over algoritmen wordt bewaard om te kunnen voldoen aan het vereiste van transparantie ten behoeve van bijvoorbeeld auditeerbaarheid, reconstrueerbaarheid en uitlegbaarheid, valt ook onder de Archiefwet. Informatie dient dan ook in haar context en in "goede, geordende en toegankelijke" staat te worden bewaard en beheerd. Om dit goed in te richten is aandacht in de design fase van belang.

[14] Passenger Name Record. Zie Wet gebruik van passagiersgegevens voor de bestrijding van terroristische en ernstige misdrijven https://wetten.overheid.nl/BWBR0042301/2019-06-18..
[15] Richtlijn (EU) 2016/680 van 27 april 2016 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door bevoegde autoriteiten met het oog op de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, en betreffende het vrije verkeer van die gegevens.
[16] Het transparantievereiste uit de AVG (artikel 5, eerste lid, onder a.) vergt dat persoonsgegevens worden verwerkt op een wijze die ten aanzien van betrokkenen transparant is. Dit beginsel wordt uitgewerkt in verschillende informatieverplichtingen. Een algemeen kader daarvoor is te vinden in artikel 12 AVG. Zie ook artikel 13 en 14 AVG.
[17] Om redenen van efficiency kan worden overwogen een BIA gelijktijdig met een DPIA uit te voeren.

Impact assessment mensenrechten en algoritmen
Naast voornoemde instrumenten en normenkaders, worden de richtlijnen versterkt middels een *impact assessment mensenrechten en algoritmen*.[18] Hiermee worden organisaties niet alleen ondersteund bij de toepassing van de richtlijnen en de vraag of zij de richtlijnen juist en voldoende hebben geïmplementeerd, maar ook de impact die (de keuze voor en de toepassing van) een algoritme heeft op de in het geding zijnde mensenrechten.

In relatie tot de DPIA kan het *impact assessment mensenrechten en algoritmen* als los instrument (en volgtijdelijk) worden uitgevoerd. In het verlengde hiervan kan worden overwogen om, bij algoritmen die gebruik maken van persoonsgegevens, de (verantwoording over de) toepassing van de richtlijnen in te bedden in de in de organisatie reeds bestaande structuur voor de uitvoering van DPIA's.

AI-systeemprincipes voor non-discriminatie
De handreiking legt uit welke vragen en principes leidend zijn bij het ontwikkelen en implementeren van een AI-systeem met het oog op het discriminatieverbod, vanuit zowel juridisch, technisch, als organisatorisch perspectief. Het document is bedoeld voor systeembouwers, data analisten en AI-experts. Dit document kan aanvullend naast de richtlijnen worden gehanteerd voor een passende borging van anti-discriminatie in AI-systemen is het noodzakelijk om een vertaling te maken van het juridisch kader naar concreet toepasbare systeemprincipes, oftewel van normen naar concrete ontwerpstrategieën die kunnen worden gebruikt bij het ontwerp van AI in een concreet geval door de overheid.

Richtlijn inzake publieksvoorlichting over data-analyses
Met deze richtlijn wordt beoogd het publiek te informeren over het gebruik van data-analyses, in het bijzonder data-analyses waarin persoonsgegevens worden verwerkt. Uit de AVG vloeit niet voort dat informatie hierover openbaar dient te worden gemaakt voor het publiek. In die zin kunnen deze richtlijnen worden gezien als een aanvulling op de transparantieverplichtingen uit de AVG die bedoeld zijn om de betrokkenen, wiens persoonsgegevens worden verwerkt, te informeren.

Transparantie over het gebruik van data-analyses kan ook worden ingegeven door de Wet openbaarheid van bestuur (Wob) en de daarin opgenomen verplichting tot openbaarmaking van informatie over bestuurlijke aangelegenheden (art. 3, 8, 10 en 11) dan wel door het zorgvuldigheids- en motiveringsbeginsel uit de Algemene wet bestuursrecht (artikelen 3:2 en 3:46 Awb). De richtlijnen kunnen worden gezien als een invulling van de openbaarmakingsverplichting uit de Wob en geven tevens invulling aan de genoemde beginselen op grond van de Awb.[19]

[18] Het Algoritme Impact Assessment wordt in het eerste kwartaal van 2021 door het ministerie van BZK aan de Tweede Kamer aangeboden.
[19]Zie ook 'Ongevraagd advies van de Raad van State over de effecten van digitalisering voor de rechtstatelijke verhoudingen', Kamerstukken II, 2017/18, 26643, nr. 557. Zie over de betekenis van de verscherpte motiveringseisen van de afdeling bestuursrechtspraak van de Raad van State, voor geautomatiseerde besluitvormingsprocessen, de antwoorden d.d. 12 juni 2019 van de minister van Rechtsbescherming op de vragen van het lid Buitenweg (Groen Links) over de motivering van automatisch genomen besluiten. Aanhangsel Kamerstukken II, 2018/19, 3088.

5. Voor welke functionarissen zijn de richtlijnen bedoeld

Richtlijn voor het toepassen van algoritmen
De inzet van algoritmische data-analyses vindt zoals gezegd plaats in een bredere bestuurlijke en organisatorische context, waarbij naast degenen die een rol vervullen bij de daadwerkelijke ontwikkeling en beheer van algoritmen, zoals softwareontwikkelaars, ontwerper/onderzoeker, beheerder en architect ook andere functionarissen betrokken zijn. Er kunnen in dat opzicht de volgende niveaus en rollen worden onderscheiden:

- een bestuurlijk niveau, waar de verantwoordelijkheid ligt voor de besluitvorming en verantwoording over de inzet van algoritmen en voor het proces;
- een beleidsniveau waar de afweging en vertaling van beleid naar uitvoering via de inzet van algoritmen plaatsvindt;
- een ontwikkel- en beheersniveau waar het ontwerpen, maken en onderhouden van algoritmen plaatsvindt;
- een controleniveau dat toetst of algoritmen juist en rechtmatig worden ingezet. Denk aan de functionaris gegevensbescherming, de interne controller of externe auditor;
- Een communicatieniveau dat met de burger communiceert over organisatieprocessen of individuele beslissingen. Dit zijn rollen die door communicatie- en beslismedewerkers worden vervuld.

Deze richtlijn richt zich primair op de ontwikkelaars, ontwerpers en beheerders, oftewel het ontwikkel- en beheersniveau, maar draagt ook bij de samenwerking met en de interne dialoog tussen de functionarissen uit de andere niveaus, over het treffen van de juiste maatregelen en waarborgen en het bepalen van de respectievelijke betrokkenheid en verantwoordelijkheid.

De mate waarin deze niveaus en bijbehorende rollen betrokken en verantwoordelijk zijn, verschilt per fase.

Het is aan organisaties om, rekening houdend met de specifieke context van de organisatie, te bepalen welke functionaris(sen) welke onderdelen van de richtlijn uitvoert. Daarbij hoort dat binnen de organisatie, de regie over het hele proces, ook na ontwikkeling en implementatie van het systeem (wanneer het systeem een lijn-activiteit is geworden) zodanig is belegd dat betekenisvolle verantwoording over het ontwerp en het gebruik van de algoritmen kan worden afgelegd.

Richtlijn inzake publieksvoorlichting over data-analyses
Deze richtlijn is primair bedoeld voor de communicatiemedewerkers, die verantwoordelijk zijn voor de informatievoorziening op de website en/of aan het publiek. Voor het bepalen van de inhoud en de mate van voorlichting zullen de (privacy)juristen en ontwikkelaars of beheerders van het algoritmische systeem betrokken moeten worden. Hierbij wordt onderscheidt gemaakt tussen algemene informatie aan het publiek (bv. via een privacy statement) en het informeren van een individuele betrokkene over het gebruik en de inzichtelijkheid van algoritmen.

6. Wie is verantwoordelijk voor het toepassen van de richtlijnen

Formeel is de betreffende minister, gedeputeerde, wethouder of bestuurder verantwoordelijk voor de taakuitoefening door (een onderdeel van) de desbetreffende overheidsorganisatie, en dus ook voor de inzet daarbij van algoritmen.
In de praktijk zal deze bevoegdheid zijn gemandateerd aan een functionaris in de lijn. De gemandateerde functionaris is dan verantwoordelijk voor de inzet van algoritmen en de toepassing daarbij van de richtlijnen en andere relevante instrumenten en normen.

Wanneer meerdere organisaties betrokken zijn bij de ontwikkeling van algoritmen, ligt het in de rede om te regelen dat één organisatie het voortouw heeft bij de ontwikkeling daarvan en de toepassing van de richtlijn en andere relevante normen en instrumenten, zoals een DPIA en IAMA. Dat laatste doet niet af aan de respectievelijke c.q. gezamenlijke verantwoordelijkheid van de betrokken organisaties (op grond van bijv. de AVG of Richtlijn).[20] De AVG verplicht gezamenlijk verwerkingsverantwoordelijken tot het opstellen van een document waarin zij neerleggen hoe de verantwoordelijkheden voor de naleving van de wet verdeeld zijn (art. 26 (1) AVG). Zie ook Deel II, Richtlijn voor het toepassen van algoritmen door overheden, onder Verantwoording.

7. Wanneer zijn de richtlijnen van toepassing

Deze richtlijnen zijn relevant in de verschillende opeenvolgende ontwikkelingsfasen van een algoritme (ontwerp, ontwikkeling, toepassing, beheer en controle).Tevens wanneer er sprake is inkoop van algoritmen, de data-analyses gebaseerd op deze algoritmen, eventuele besluitvorming op basis van de analyses, en de evaluaties ervan. Deze zijn daarbij relevant voor alle types algoritmen.[21] Deze richtlijnen zijn ook van toepassing indien een systeem/ algoritmen door een externe partij wordt ontwikkeld en door de overheid wordt aangeschaft. Ook in dat geval dient het algoritme aan de richtlijnen te voldoen.

Aanmerkelijke impact
Het is vooraf niet *exact* te bepalen op welke specifieke algoritmen de richtlijnen van toepassing zijn en in hoeverre algoritmen daaraan moeten voldoen, want dit is per geval verschillend. Wel kan in algemene zin gesteld worden dat de richtlijnen van toepassing zijn op data-analyses die rechtsgevolgen of anderszins een aanmerkelijke impact hebben op burgers, bedrijven, of (groepen in) de samenleving. De impact van de data-analyse op de burger, bedrijven of samenleving is met andere woorden bepalend. Hoe groter de impact en autonomie van het algoritme, hoe stringenter de richtlijnen zullen moeten worden toegepast. Overigens hoeft het niet alleen om besluitvorming door de overheid te gaan. Ook algoritmen die worden gebruikt om risico's op bijvoorbeeld fraude in te schatten of ter onderbouwing van beleid kunnen een aanmerkelijke impact op burgers hebben. Om die reden wordt niet enkel gesproken over besluiten maar ook over 'uitkomsten'.

De richtlijnen onderscheiden algoritmen op de assen van complexiteit, autonomie en impact. Naast eisen die voorkomen uit wetgeving omtrent autonome gegevensverwerking, is ook de mate van menselijke controle hiervoor anders. Dit leidt bij implementatie van de richtlijn tot een groter belang omtrent validatie en transparantie op basis van deze drie dimensies:

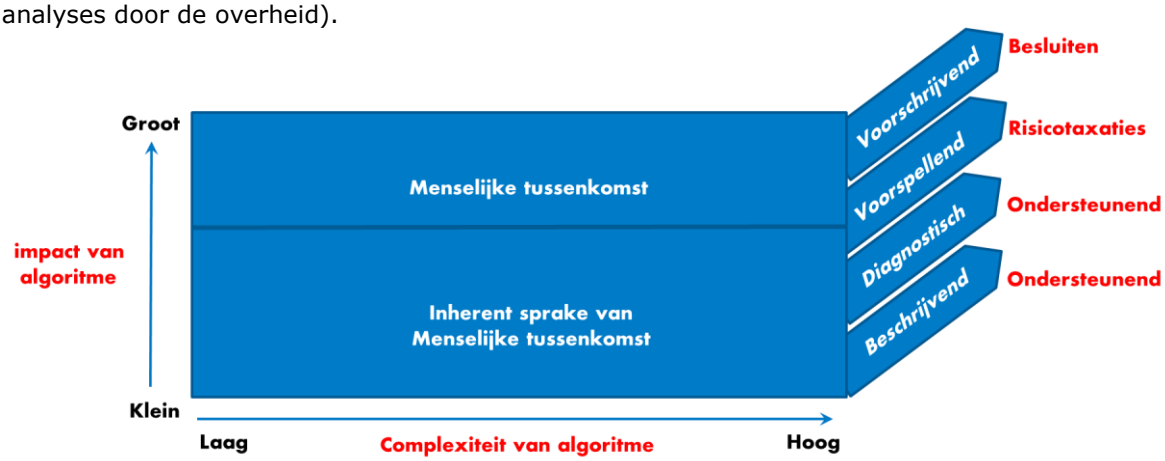
- Complexiteit
- Autonomie
- Impact

Een algoritme dat laag scoort op impact, autonomie en complexiteit kent een andere mate van naleving van de richtlijnen dan algoritmen die hoog scoren op al deze drie dimensies.

[20] Zie over de gezamenlijke verantwoordelijkheid van verwerkingsverantwoordelijken, artikel 26 AVG en 30 Richtlijn.
[21] De richtsnoeren van de HLEG AI focussen daarentegen op de algoritmen die onder de categorie Kunstmatige Intelligentie vallen.

Complexiteit
Uitgangspunt is dat overheidsorganisaties in beginsel geen algoritmen hanteren die te complex zijn om redelijkerwijs te kunnen worden uitgelegd.

Autonomie
Volledigheidshalve, is in onderstaande figuur getracht duidelijk te maken dat daar waar de inzetgebieden beschrijvend, diagnostisch en voorspellend zijn er inherent sprake is van menselijke tussenkomst, dat wil zeggen dat op de uitkomst van het algoritme *altijd* een menselijke handeling volgt. Daar waar het inzetgebied voorschrijvend en dus besluitvormend is, en er voorts sprake is van een situatie als bedoeld in artikel 22 AVG, geldt dat het recht op (betekenisvolle) menselijke tussenkomst moet worden georganiseerd (artikel 22 AVG). Zie hierover voornoemde brief van 8 oktober 2019, blz 9-10. (waarborgen tegen risico's van data-analyses door de overheid).



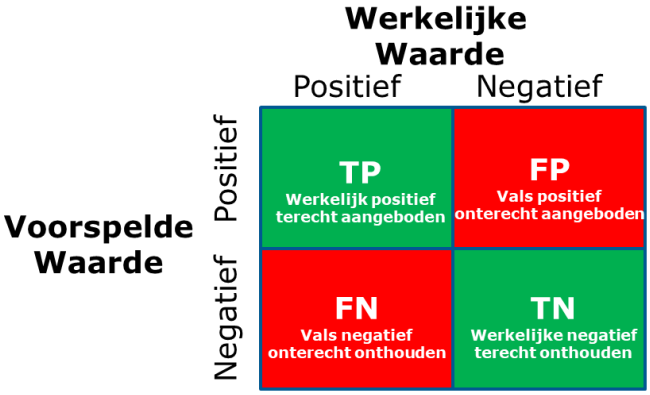
Impact
Voor het bepalen van de impact dient te worden gekeken naar:
1. Wat is de impact van het algoritme op de beslissing en uitkomst?
   Met andere woorden hoe groot is de invloed en autonomie van het gebruikte algoritme (data-analyse) in het totaal van de context waarvoor het wordt ingezet, het daarbij behorende proces en de uitkomst in de vorm van bijvoorbeeld een beslissing, risicotaxatie of beleid?
2. Wat is de impact van de data-analyse en de uitkomsten daarvan op de burger, bedrijf of samenleving?

Ad 1. Impact van het algoritme op de (uitkomst van de) data-analyse
Naarmate de betekenisvolle menselijke tussenkomst tijdens een data-analyse beperkter is, is de impact van het algoritme op de uitkomst groter. Dat stelt hogere eisen aan procesinrichting en –bewaking, en de mogelijkheden om processen stil te kunnen leggen, terug te draaien en achteraf te kunnen toetsen en controleren (via audits en eisen inzake toetsbaarheid en uitlegbaarheid). En daarmee ook aan de toepassing van de richtlijnen.

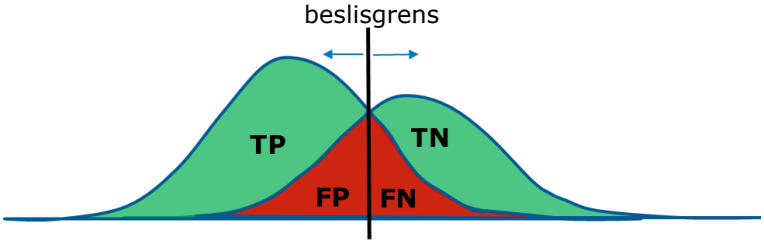
Ad 2. Impact van de (uitkomst van de data-analyse) op burger, bedrijf of samenleving
Gelet op het benodigde vertrouwen van burgers in het gebruik van algoritmen door de overheid is ook de impact van de uitkomst van de data-analyse op de burger relevant.

Voorbeeld is een casus van het UWV over de werkverkenner. De werkverkenner bepaalt geautomatiseerd wanneer een WW'er een persoonlijk gesprek krijgt aangeboden. Het gebruikte model kan er in de voorspelling naast zitten, waardoor iemand onterecht een gesprek wordt onthouden of aangeboden. Als de voorspelling van het algoritme ervoor zorgt dat een gesprek *"onterecht" aangeboden* wordt, is de negatieve impact beperkt tot de energie en tijd die van beide kanten aan het gesprek wordt besteed. Als de voorspelling van het algoritme ervoor zorgt dat een gesprek 'onterecht' onthouden wordt, is de mogelijke negatieve impact veel groter: het ondersteunen van werkzoekenden in de eerste maanden is volgens onderzoek namelijk vele malen effectiever. Deze mogelijke negatieve impact stelt eisen aan het geautomatiseerd proces waarin voor dit geval vangnetten moeten worden georganiseerd.

Het bovenstaande dilemma wordt ook wel de verwarringsmatrix (confusion matrix) of foutenmatrix (error matrix) genoemd, hieronder afgebeeld. Waarbij gebaseerd op bovenstaand voorbeeld op basis van voorspelende waarde een gesprek wordt aangeboden of onthouden.



Hierdoor ontstaat de afwegingskeuze waar de zogenaamde beslisgrens moet liggen.
Met de extremen, geen enkel gesprek onterecht onthouden, met als gevolg vele onterecht aangeboden. Of geen enkele onterecht aangeboden met als gevolg vele onterecht onthouden. Deze afweging waar de beslisgrens ligt moet van tevoren duidelijk worden gewogen en worden bepaald.



Bij data-analyses middels algoritmen die in besluitvorming uitmonden zal in de regel sprake zijn van (rechtsgevolgen[22] en dus van) aanmerkelijke impact op de burger of bedrijf. Zoals eerder aangegeven is er bij voorschrijvende algoritmen veelal sprake van besluitvorming. Dit stelt hogere eisen aan de transparantie van deze processen (in termen van uitlegbaarheid, toetsbaarheid, auditeerbaarheid) en zal veelal vereisen dat betekenisvolle menselijke tussenkomst wordt georganiseerd om risico's voor burgers te beperken. Dat laatste hoeft overigens niet altijd het geval te zijn. Een voorbeeld betreft de administratiefrechtelijke afdoening van verkeersovertredingen onder de wet administratiefrechtelijke handhaving verkeersvoorschriften (Wet Mulder).[23] Hierbij is geen sprake van menselijke tussenkomst omdat de uitkomst volledig op regelgeving gebaseerd is.

[22] In de Ethische richtsnoeren is vermeld dat de drempel voor "aanmerkelijke mate" vergelijkbaar moet zijn met de mate waarin de betrokkene wordt getroffen bij een besluit waaraan een rechtsgevolg verbonden is. Andersom betekent dit dat er bij algoritmen die uitmonden in besluitvorming waaraan rechtsgevolgen zijn verbonden, in de regel sprake zal zijn van aanmerkelijke impact. Zie hierover Rechtbank Den Haag, 5 februari 2020, ECLI:NL: RBDHA:2020:865, r.o. 6.36.

[23] Omdat het om gebonden bevoegdheden gaat waarbij de algoritmische besluitvorming plaatsvindt op basis van vooraf en objectief vastgestelde variabelen die in een directe relatie tot de betrokken persoon staan, heeft het geen

Ook bij voorspellende algoritmen zal veelal sprake zijn van aanmerkelijke impact op betrokkenen. Tot deze conclusie kwam de rechtbank Den Haag in de zaak betreffende het risicotaxatiemodel SyRi. Volgens de rechtbank heeft een risicomelding een aanmerkelijke impact op het privéleven van degene op wie de melding betrekking heeft. "Een risicomelding kan voor gedurende twee jaar worden opgeslagen en mag voor (maximaal) twintig maanden door de deelnemers van het desbetreffende SyRi-project worden gebruikt. Verder mag ook aan het Openbaar Ministerie en aan de politie desgevraagd mededeling van de risicomelding worden gedaan. Dat een risicomelding niet steeds tot verder onderzoek hoeft te leiden, of tot een (bestuursrechtelijke of strafrechtelijke) sanctie en ook niet als enige basis voor een handhavingsbesluit mag worden gebruikt, doet niet af aan het aanmerkelijke effect op het privéleven van een betrokkene".[24]

De Ethische Richtsnoeren voor betrouwbare Kunstmatige Intelligentie bieden een behulpzaam criterium voor het bepalen of een algoritmische data-analyse een aanmerkelijke impact heeft op een burger, bedrijf of (groep in de) samenleving: 'Gegevensverwerking treft iemand volgens de richtsnoeren in aanmerkelijke mate wanneer de effecten van de verwerking groot of belangrijk genoeg zijn om aandacht te verdienen. Het besluit moet het potentieel hebben om de omstandigheden, het gedrag of de keuzen van de betrokken personen in aanmerkelijke mate te treffen, een langdurig of blijvend effect op de betrokkene te hebben, of in het uiterste geval, tot uitsluiting of discriminatie van personen te leiden'.[25]

Impact Assessment
Ter uitvoering van bovenstaande toetsingen zal een beoordeling moeten worden gemaakt van de (te verwachten) impact/gevolgen (en bijbehorende risico's) van de (voorgenomen) algoritmische data-analyse. Het uitvoeren van deze toetsingen vergt met andere woorden, dat een impact assessment wordt uitgevoerd. Hiervoor kan gebruikt worden gemaakt van eerdergenoemde *Impact Assessment mensenrechten en algoritmen.*

toegevoegde waarde als in dit proces een vorm van menselijke tussenkomst zou plaatsvinden. Hetzelfde geldt voor het toekennen van kinderbijslag of het bijstellen van de hoogte van het recht op studiefinanciering op basis van veranderingen in het inkomen van een van de ouders. Zie Kamerstukken II, 2017/2018, 34851, nr. 7, blz 72-73.
[24] Zie Rechtbank Den Haag, 5 februari 2020, ECLI:NL: RBDHA:2020:865, r.o. 6.59.
[25] Op dit criterium baseerde de Rechtbank Den Haag zich voor haar oordeel in de zaak SyRi dat een risicomelding een aanmerkelijke impact heeft op betrokkenen.
Kanttekening: de rechtbank spreekt zich niet uit over de vraag of de inzet van SyRi als individuele geautomatiseerde besluitvorming in de zin van artikel 22 AVG moet worden gekwalificeerd. De rechter beperkt zich tot het oordeel dat de risicomelding een 'aanmerkelijke effect' heeft op betrokkenen.

8. Monitoring en doorontwikkeling van de richtlijnen

De technologie met betrekking tot data-analyses is voortdurend in ontwikkeling en zal dat naar verwachting ook blijven. Daarnaast zijn er, op zowel nationaal, Europees als internationaal niveau, diverse ontwikkelingen en initiatieven gaande met betrekking tot het normeren en reguleren van algoritmen en meer specifiek van AI. Naast de technologische ontwikkelingen zijn ook deze ontwikkelingen relevant voor de richtlijnen, de kwaliteit en effectiviteit daarvan, en maken dat het nodig is om de richtlijnen periodiek te evalueren en te blijven door ontwikkelen.

De richtlijnen zullen in brede afstemming met andere overheidspartijen periodiek worden geëvalueerd en aangepast op nieuwe inzichten of methoden. Het zal worden onderzocht hoe dit het beste kan worden vormgegeven.

Deel II – de richtlijnen

Richtlijn voor het toepassen van algoritmen door overheden

1. Bewustzijn en inperking van risico's

Systeemarchitecten, systeemontwerpers, onderzoekers, ontwikkelaars, beheerders, en gebruikers van algoritmische systemen, die gebruikt worden voor toepassingen met consequenties voor burgers of groepen in de samenleving (zie, Deel I, paragraaf 7), moeten zich bewust zijn van, en maatregelen treffen om de risico's op fouten en ongewenste bias die data-analyses en daarbij gebruikte datasets, algoritmen of methoden in zich hebben, te beperken zodanig dat de toepassing ervan in overeenstemming is met de wet. Hetzelfde geldt voor de mogelijke discriminerende- of stigmatiserende effecten en factoren die het ontwerp, de implementatie daarvan en het gebruik kunnen opleveren.

Met als specifieke aandachtspunten en maatregelen:

- Test het algoritme op basis van test cases of scenario's en evalueer test cases periodiek en elke keer als de software verandert om te voorkomen dat nieuwe fouten ontstaan dan wel functionaliteit onbedoeld wordt aangepast. Creëer dus zogenaamde feedbackloops.
    - □ Test het algoritme op basis van test cases/scenario's.
    - □ Evalueer de test cases/scenario's periodiek.
    - □ Evalueer de test cases/scenario's wanneer de software is veranderd.

- Houd rekening met het feit dat een verandering of af- of toename van de gegevens in de tijd van invloed kan zijn op de uitkomsten van een algoritme en corrigeer daar zo nodig op.
    - □ Monitor het algoritme wanneer er sprake is van 1 of meerdere van de onderstaande punten:
        1. Verandering van data
        2. Afname van de data
        3. Toename van de data
        4. Verandering van de dataverdeling
    - □ Test en pas het algoritme aan indien nodig wanneer uitkomsten van het algoritme zijn veranderd op een manier dat het doel van het algoritme niet meer ondersteund wordt of wanneer er ongewenste biases in de uitkomst ontstaan.

- Zorg, binnen de grenzen van de wetgeving[26] voor het opbouwen van controlemechanismes die specifiek toetsen of er geen sprake is van discriminatie of stigmatisering.
    - □ Test binnen de grenzen van de wetgeving of er sprake is van discriminatie of stigmatisering door het algoritme. Betrek hierbij de AI-systeemprincipes ter bestrijding van discriminatie.
    - □ Maak inzichtelijk wat het gewicht (de feature importance) is van de voorspellende variabelen op de uitkomsten van het model.
    - □ Maak inzichtelijk hoe de relatie tussen iedere voorspellende variabele en modeluitkomst er uit ziet. Is de relatie bijvoorbeeld lineair of non-lineair, en wel of niet monotoon?
    - □ Onderzoek of bepaalde voorspellende variabelen zoals buurt of wijk proxy's zijn voor (bijzondere) persoonsgegevens en leiden tot ongewenste discriminatie.

[26] Soms kan het nodig zijn om binnen de test-set te werken met zgn. bijzondere persoonsgegevens om te kunnen constateren of sprake is van discriminerende of stigmatiserende effecten. Echter, de vigerende wetgeving (artikel 9, AVG en artikel 22 – 30 Uitvoeringswet AVG) laat controle met behulp van bijzondere persoonsgegevens, zoals gegevens over iemands ras, vooralsnog niet toe. Hiervoor is aanpassing van de AVG of UAVG nodig. Zowel op nationaal niveau als in Europees verband heeft het kabinet laten weten voorstander te zijn van een dergelijke wetswijziging. Zie kabinetsbrief van 9 oktober 2019 over waarborgen tegen risico's van data-analyses door de overheid (Kamerstukken II 2019/20, 26643, nr. 641) en Kabinetsappreciatie Witboek over Kunstmatige intelligentie (Kamerstukken II, 2019/20, 26643, nr. 680, p. 12).
Discriminatie is verboden, en dus onrechtmatig, wanneer sprake is van een verboden onderscheid. Daarvan is sprake wanneer onderscheid wordt gemaakt op grond van beschermde persoonskenmerken uit de gelijke behandelingswetgeving, zoals bijvoorbeeld ras, seksualiteit of politieke overtuiging. Het belang van het voorkomen van discriminerende factoren of effecten van algoritmen volgt uit artikel 5 AVG dat bepaalt dat persoonsgegevens rechtmatig moeten worden verwerkt.

- Onderzoek de kwaliteit van databronnen en is deze voldoende voor het doel waarvoor deze worden ingezet. Breng in kaart of/welke beperkingen het gebruik van een databron, algoritme of analysemethode kent, bijvoorbeeld doordat er rechten zijn gevestigd op de databron, algoritme of methode. Een andere beperking is dat een databron waarvan de kwaliteit van de data slechts globaal of veranderlijk is, minder geschikt kan zijn voor data-analyses die leiden tot individuele voorspellingen of beslissingen. Een dergelijke bron kan wel geschikt zijn voor het beschrijven van groepen.

  Datakwaliteit is essentieel voor het kunnen uitvoeren van een gedegen analyse en brongegevens bevatten eigenlijk altijd biases en fouten. Datakwaliteit is des te meer van belang wanneer gebruik wordt gemaakt van persoonsgegevens, waarvoor volgens de AVG en de Richtlijn geldt dat die juist zijn en zo nodig worden geactualiseerd.[27] Ook dienen de gegevens noodzakelijk te zijn voor het doel van de analyse.[28] Beschrijf daarom ook het doel van de analyse en hoe je omgaat met de data en fouten in de datasets en uitkomsten.
  - □ Beschrijf het doel van de analyse
  - □ Onderzoek of de data(bron) van voldoende kwaliteit & kwantiteit is[29]:
    1. Is de data relevant voor het doel?
    2. Is de data-aggregatie op een niveau voldoende informatief voor het doel.
    3. Wordt de data over tijd consistent bijgehouden?
    4. Wordt de data op dezelfde manier geregistreerd?
    5. Is er voldoende variatie binnen de data? Zijn bepaalde deelpopulaties (minderheden) oververtegenwoordigd in de data, wat zou kunnen leiden tot bias in het voorspellend model?
    6. Is de gebruikte dataset juist, tijdig en volledig?
    7. Zijn de databronnen te koppelen (indien er wordt gewerkt met meerdere tabellen)?
    8. Zijn eventuele mutaties in de data(bron) te herleiden?
    9. Is er een database wijziging geweest?
  - □ Onderzoek de beperkingen:
    1. Zijn er rechten gevestigd op de databron?

- Maak een bewuste keuze voor data-analyse technieken. Het is bijvoorbeeld niet zonder meer noodzakelijk kunstmatige intelligentie methoden in te zetten op data. Vaak zijn ook andere analysemethoden geschikt om de kwaliteit van bronnen te onderzoeken of om patronen te vinden. Belangrijk is ook of je wilt werken met vooraf bedachte hypothesen die je wilt toetsen d.m.v. data-analyse, of dat je zonder hypothese te werk wilt gaan. In dat laatste geval is het in het algemeen lastiger om een werkwijze te legitimeren.
  - □ Welke technieken zijn getest?
  - □ Is er een minder verstrekkende techniek getest?
  - □ Waarom is er gekozen om deze specifieke technieken te testen?
  - □ Waarom is er voor een bepaalde techniek of een combinatie van technieken gekozen?
  - □ Indien er uiteindelijk voor een kunstmatige intelligentie techniek is gekozen, licht toe waarom.
  - □ Is er gewerkt met vooraf bedachte hypothesen?

- Hanteer een standaard methodiek van werken, zoals CRISP DM (Cross Industry Standard Process for Data Mining). Deze methodiek is toepasbaar bij datamining (techniek om patronen in datasets te ontdekken). Wanneer je een hypothese wil toetsen aan de hand van data-analyse is deze methodiek minder van toepassing.
  - □ Welke standaard methodiek wordt gebruikt? Beschrijf stappen van de modelontwikkeling. Bijvoorbeeld:
    - De data
      - Brondata en typen data
      - Biases & correctie
      - Data-kwaliteit & kwantiteit

[27] Artikel 5, eerste lid, onder d, AVG. Zie ook Afdeling 3 Rectificatie en wissing van gegevens, AVG.
[28] Wanneer met persoonsgegevens wordt gewerkt schrijft de AVG voor dat de gegevensverwerkingen worden beperkt tot wat noodzakelijk is voor de doeleinden waarvoor zij worden verwerkt (data-minimalisatie). De gegevens moet m.a.w worden beperkt, maar ook voldoende zijn voor het doel van de data-analyse. Zie artikel 5, eerste lid, onder c, AVG.
[29] De focus hier is het onderzoek naar de kwaliteit van de data(bronnen) en hoe men daarmee omgaat. Onder Uitlegbaarheid worden maatregelen beschreven die betrekking hebben op de beschrijving van de kwaliteit van data(bronnen).

        - Omgang met gevoelige gegevens
    - Data verwerken
        - Belangrijke definities
        - Data preparatie en opschoning
        - Brontabellen werkbaar maken
        - Samenstelling & opbouw van modeltests
        - Basis dataframe
        - Variabelen creatie
        - Data fouten corrigeren
    - De modelontwikkeling
        - Aanpak modelkeuze
        - Uitleg over toegepaste modellen en technieken
    - Resultaten
        - De modelbeoordeling
        - Hoe werkt de modelbeoordeling
    - Gebruik van het algoritme in de praktijk

- Hanteer gelet op de aard van gegevensverwerking juridische en beleidsmatige toetsingskaders en zorg bij verwerking van persoonsgegevens voor de uitvoering van een DPIA om risico's voor de gegevensbeschermingsrechten van betrokkenen zoveel mogelijk te beperken.

- In het algemeen zal gelden dat bij algoritmen die gebruikt worden voor geautomatiseerde besluitvorming (voorschrijvend), gebruik gemaakt wordt van causaliteit. Dat impliceert dat in dat geval het gebruik van *deep learning* minder voor de hand ligt, omdat die techniek in toenemende mate gebruik maakt van correlaties, d.w.z. statistische verbanden. Houd m.a.w. rekening met het risico dat zelflerende algoritmen gebruik maken van correlaties en statistische verbanden waarbij het de vraag is of deze ook geschikt zijn om besluitvorming op te baseren.
    - □ Is een algoritme dat uitgaat van causaliteit noodzakelijk om het doel te bereiken?
    - □ Zo ja/zo nee, waarom?

- Hanteer na implementatie en inrichting van het algoritme een evaluatiecyclus/feedbackloop, zodat wanneer ontwerpers en architecten op afstand komen te staan risico's op tijd worden vastgesteld en adequaat geadresseerd.

- Onderzoek de beperkingen:
    1. Zijn er rechten gevestigd op het gebruik van het algoritme?
    2. Zijn er rechten gevestigd op het gebruik van de analysemethode?

  Het gegeven dat er rechten rusten op de content mag geen beperking opleveren in het inzicht dat moet worden verkregen in het algoritme.

Voor (supervised) machine learning
- Maak eisen ten aanzien van toepasbaarheid, voorspelkracht, uitlegbaarheid, ongewenste discriminatie, enz. expliciet (kwantitatief), zodat de oplossingsruimte verder verkleind wordt:
    - □ Geef aan voor welke populatie en onder welke omstandigheden het model van toepassing moet zijn.
    - □ Bepaal samen met de materie deskundigen hoe de voorspelkracht van het model beoordeeld gaat worden. Met welke metrics en bijbehorende drempelwaarden? Wanneer is het model goed genoeg? Zijn false positives bijvoorbeeld minder erg dan false negatives, of is het beide even belangrijk? Hoe kunnen we dit kwantificeren?
    - □ Bepaal samen met de business wanneer er sprake is van ongewenste discriminatie. Welke metrics en drempelwaarden gaan we gebruiken om dit meetbaar en objectiveerbaar te maken?
    - □ Bepaal samen met de business of het model gewenste bias moet vertonen. Probeer ook dit te kwantificeren.

2. Transparantie en uitlegbaarheid

Overheden die gebruik maken van algoritmische data-analyses met aanmerkelijke impact voor individuele burgers, bedrijven of (groepen in) de samenleving moeten in begrijpelijke taal uitleg kunnen geven over de uitkomsten daarvan en hoe deze tot stand zijn gekomen. Voor uitlegbaarheid is van belang dat duidelijkheid en uitleg kan worden gegeven over: het doel dat met het algoritme wordt nagestreefd, de procedures die door het algoritme worden gevolgd, het toegepaste model en algoritme(n), welke variabelen of beoordelingscriteria doorslaggevend zijn geweest voor de uitkomst en de data die worden gebruikt (de kwaliteit, herkomst en eventuele combinatie daarvan, hoe de gegevens zijn getoetst).[30]
Deze informatie is bedoeld om in concrete gevallen te worden gegeven[31]. Bijvoorbeeld op verzoek van de betrokkene, de rechter, toezichthouder of interne/externe controller.
Dit impliceert een gedegen documentatie waarin voornoemde aspecten uit het ontwikkelingsproces zijn vastgelegd. De inzet van het algoritme dient steeds proportioneel te zijn aan het doel dat ermee wordt gediend en de mogelijke inbreuken die ontstaan als gevolg van het gebruik ervan Ook brengt dit als uitgangspunt mee dat overheidsorganisaties in beginsel geen algoritmen mogen hanteren die te complex zijn om redelijkerwijs te kunnen worden uitgelegd.

Uitlegbaarheid vereist dat een organisatie intern zicht en regie heeft op het ontwikkelingsproces en dat collega's, teams, afdelingen betrokken zijn bij het ontwikkelingsproces en daarbij aan elkaar uitleggen wat zij doen en welke keuzes zij maken. Dit betreft de collegiale uitlegbaarheid.

Uitlegbaarheid leidt tot transparantie van het algoritmische proces en kan in die zin worden gezien als een nadere invulling - toegesneden op de specifieke kenmerken van algoritmische data-analyses - van het transparantiebeginsel uit de AVG[32] en de beginselen uit de Awb, zoals het zorgvuldigheidsbeginsel en het motiveringsbeginsel.

Met als specifieke maatregelen:

- Organiseer de code in modules welke separaat en gecombineerd kunnen worden geëvalueerd.
    - □ Er is een 'Main' script waarin alle codes worden aangeroepen. Per gegevensbron is er bijvoorbeeld een code waarin de data is geprepareerd.
    - □ De code is in modules georganiseerd.
    - □ Elke module kan separaat en gecombineerd worden geëvalueerd.

- Test deze modules op correcte functionaliteit zowel afzonderlijk als in combinatie.
    - □ Alle scripts zijn, waar mogelijk, afzonderlijk van elkaar getest. Alle scripts zijn in combinatie met elkaar getest. Dit is standaard onderdeel van de modelontwikkeling
    - □ De modules kunnen op correcte functionaliteit zowel afzonderlijk als in combinatie getest worden.

- Leg de gehanteerde analysemethode uit en meet de nauwkeurigheid.
    - □ In de technische documentatie is beschreven welke methode(s) is/zijn gebruikt
    - □ De gehanteerde analysemethode is uitgelegd.
    - □ De nauwkeurigheid van de analysemethode is gemeten en beschreven.

- Leg de input gegevens (brondata/datasets) vast die gebruikt worden en gebruik daarbij enkel noodzakelijke data. Documenteer en leg dit vast.
    - □ In de technische documentatie is vastgelegd welke databronnen zijn gebruikt.

[30] Transparantie over *qualifiers: m*et het oog op de uitlegbaarheid van data-analyses zijn vooral de *qualifiers* (variabelen en drempelwaarden) binnen een algoritme van belang. Welke *qualifiers* zorgen ervoor dat men tot een risicoprofiel komt, en kunnen deze *qualifiers* inzichtelijk worden gemaakt? Een toetsingscommissie zou aan de hand van *case studies* kunnen toetsen wanneer diensten over deze *qualifiers* wel en niet transparant kunnen worden gemaakt. Daarbij zou kunnen worden overwogen om, indien een risicoprofiel invloed heeft op de rechten en plichten van iemand, deze *qualifiers* voor hem of haar transparant te maken. Deze vorm van transparantie zou dan om *gaming the system* te voorkomen bij voorkeur niet vooraf in het proces moeten plaatsvinden maar achteraf.
[31] Dit is anders voor de informatie die volgens de Richtlijnen inzake publieksvoorlichting over data-analyses, aan het publiek wordt verschaft. Deze informatie wordt pro-actief gegeven en is meer algemeen van aard is. De informatievoorziening aan het publiek kan ook worden aangemerkt als 'publieke uitlegbaarheid'.
[32] Voor zover het gaat om algoritmen waarbij persoonsgegevens worden gebruikt. Zie artikel 5, eerste lid, onder a, AVG: persoonsgegevens worden verwerkt op een wijze die ten aanzien van betrokkenen transparant is. Zie ook artikel 12 t/m 14 AVG.

- ☐ De input gegevens (brondata/datasets) die gebruikt worden, zijn vastgelegd.
- ☐ De gebruikte data zijn relevant.
- ☐ De gebruikte data en de relevantie ervan is gedocumenteerd.

- Beschrijf de kwaliteit van de gebruikte databron(nen) en of de databron(nen) van voldoende kwaliteit is voor het doel waarvoor deze wordt ingezet.
  In een 'Data Deep Dive' en ook bij de eerdere inperking van risico's is de kwaliteit van de databronnen grondig onderzocht en toegelicht. Ook in de technische documentatie is beschreven hoe de data-kwaliteit is onderzocht. Alleen databronnen met voldoende kwaliteit worden meegenomen in het model.
  - ☐ De kwaliteit van de gebruikte databron(nen) is onderzocht.
  - ☐ De kwaliteit van de gebruikte databron(nen) is beschreven.

- Leg de aannames/keuzes die gehanteerd zijn vast. Het gaat hier niet om alle aannames tijdens het programmeren inzichtelijk te maken, maar om de keuzes die zijn gemaakt: bijv. om bepaalde data niet mee te nemen in de analyse omdat de kwaliteit daarvan onvoldoende is of omdat een dataset onvoldoende gegevens bevat om een statistische analyse op uit te voeren.
  - ☐ Alle keuzes zijn beschreven in de technische documentatie.
  - ☐ Alle keuzes zijn gedocumenteerd

Collegiale uitlegbaarheid: Zorg ervoor dat teams volledig toegang/inzicht hebben in elkaars documentatie, beslissingen en code. Wanneer beslissingen over features, specificaties, ontwerp, bouw en tests verdeeld zijn over meerdere teams kunnen er in de overdracht ongemerkt en onbedoeld interpretatieverschillen ontstaan. Transparantie en uitlegbaarheid komen dan in gevaar.

Waar de bewaarde informatie "uitlegbaarheid" dient, zal deze veelal de vorm hebben van "reguliere" stukken, zoals Word-of PDF-documenten of (digitale) presentaties.
waarbij ook deze documentatie moet voldoen aan de vereisten van duurzame toegankelijkheid.
Waar het gaat om technische transparantie zal de te bewaren informatie eerder de vorm hebben van (bron)code en data. Voor deze informatie is de kans groter dat deze wordt beheerd in andere omgevingen, zoals *code repositories* en databases. Extra aandacht dient besteed te worden aan het zorgen dat ook deze informatie in beeld is bij de verantwoordelijken voor informatiebeheer zodat ook deze informatie toegankelijk is en voldoet aan o.a. de Archiefwet.

3. Gegevensherkenning (non-discriminatie, ongewenst profileren)

Wanneer gebruik gemaakt wordt van methoden waarbij vooraf parameters moeten worden vastgesteld of trainingsgegevens worden gebruikt, beschrijf dan de wijze waarop de parameterisering en de keuze voor trainingsgegevens tot stand is gekomen, vergezeld van een verkenning van de potentiële discriminerende factoren. Om onbedoelde en ongewenste effecten tegen te gaan kan het juist nodig zijn om bij de ontwikkeling van modellen en methoden relevante bijzondere persoonsgegevens te verwerken.[33] Zo kunnen elementen in het profileringsproces die tot vooroordelen kunnen leiden, worden geëlimineerd en kunnen verkapte tekortkomingen in datasets of modellen worden gecorrigeerd. In een recent experimenteel onderzoek is een model ontwikkeld waarmee aangetoond is dat door gebruik van relevante bijzondere persoonsgegevens als controle-variabele, discriminerende effecten gecorrigeerd kunnen worden.[34] Daarom valt te overwegen om bij de ontwikkeling van modellen en, zo nodig, de uitvoering van maatregelen, toe te staan dat bijzondere categorieën van persoonsgegevens worden verwerkt, voor zover dat noodzakelijk is om discriminerende effecten tegen te gaan.

Leg de trainingsgegevens of andere gebruikte informatie om te komen tot parameterisering vast, zodat het mogelijk is resultaten te reproduceren.

[33] Zie in die zin ook: H. Lammerant, P. Blok & P. de Hert, Big data besluitvormingsprocessen en sluipwegen van discriminatie, NTM/NJCM-bull. 2018/1, p. 10-11.
[34] Zie hierover: M. van der Sangen, Onderzoek naar eerlijke algoritmes voor beleid, 6-8-2019, https://www.cbs.nl/nl-nl/corporate/2019/32/onderzoek-naar-eerlijke-algoritmen-voor-beleid.

Maak, indien mogelijk, analyses met betrekking tot de gevolgen die een andere keuze van parameterisering of inzet van trainingsdata (meer of minder, volgorde van aanbieden van trainingsdata) heeft op de resultaten.

- □ Beschrijf de wijze waarop parameterisering tot stand is gekomen.
- □ Beschrijf de wijze waarop de keuze voor trainingsgegevens tot stand is gekomen.
- □ Verken potentiele discriminerende factoren en documenteer deze.
- □ Leg vast welke trainingsgegevens gebruikt zijn.
- □ Leg vast welke parameters gebruikt zijn.
- □ Analyseer en beschrijf wat de gevolgen zijn van een andere parameter keuze voor de resultaten.
- □ Analyseer en beschrijf wat de gevolgen zijn van een andere trainingsdataset voor de resultaten.
- □ Het bovenstaande wordt in de technische documentatie beschreven.

Om het risico op discriminatie te voorkomen die te worden onderzocht:

- □ Hoe wordt er omgegaan met bias in de data?
- □ Hoe wordt er omgegaan met fouten in de data?
- □ Hoe wordt er omgegaan met outliers in de data?
- □ Hoe wordt er omgegaan met missende waarden?

Maak onderscheid tussen gegevens die worden gebruikt:

- □ Om een model te trainen of te ontwikkelen.
- □ Die als onderdeel van het proces worden verzameld/ingevoerd. Denk hierbij aan data die nodig is als input om een Pilot uit te voeren.
- □ Die worden opgevraagd uit andere processen of anderen bronnen. Denk hierbij aan data die nodig is om analyses uit te voeren om na te gaan of de pilot een succes is.

4. Auditeerbaarheid

Auditeerbaarheid richt zich met name op het proces. Modellen, algoritmen, data en beslissingen met aanmerkelijk effect voor individuele burgers, bedrijven of groepen in de samenleving moeten worden gedocumenteerd en vastgelegd, zodat ze achteraf geverifieerd kunnen worden. Dit betekent een gedegen R&D-proces plus documentatie waarin het gebruik van algoritmen in productie navolgbaar is.

Met als specifieke aandachtspunten en maatregelen:

- Algoritmen dienen zo te worden ontworpen en beheerd dat deze toegankelijk zijn voor controllers en toezichthouders en bij voorkeur ook voor experts en burgers. Dat impliceert het volgende:
  - het algoritme dient niet-confidentieel te zijn;
    - □ Het algoritme is niet-confidentieel (non-proprietary). De werking is inzichtelijk en kan ge-audit worden.
    - □ Indien het algoritme niet wordt gepubliceerd, moet het ge-audit kunnen worden.
  - het algoritme dient gedocumenteerd te zijn;
    - □ Het algoritme is gedocumenteerd, d.w.z., bij de code is beschreven wat er gebeurt en er is een technische documentatie beschikbaar waarin de modelontwikkeling is beschreven.
  - gebruik zoveel mogelijk algoritmen en analysemethoden die wetenschappelijk gevalideerd zijn;
    - □ Er zijn uitsluitend algoritmen gebruikt die wetenschappelijk gevalideerd zijn.
    - □ De algoritmen/analysemethoden zijn (zoveel mogelijk) wetenschappelijk gevalideerd.
  - gebruik indien mogelijk algoritmen die reeds open source zijn of stel deze als open source beschikbaar.
    - □ Het algoritme is open source (indien mogelijk).
    - □ Er zijn alleen gepubliceerde (wiskundige) methoden en technieken gebruikt (zoals Random Forest).
    - □ Publiceer het algoritme indien mogelijk

Er kunnen redenen zijn om van bovenstaande richtlijnen op basis van een juiste onderbouwing af te wijken. Echter aan het uitgangspunt dat het algoritme navolgbaar en controleerbaar dient te zijn, dient ook in dat geval te worden voldaan.

- Onderbouw keuzes, zoals de keuze voor specifieke algoritmen en gebruikte data.
    - □ In de technische documentatie is beschreven waarom het uiteindelijke algoritme gekozen is en welke data gebruikt zijn.
    - □ De keuze voor de gebruikte algoritmen is onderbouwd.
    - □ De keuze voor de gebruikte data is onderbouwd.

- Noteer waarnemingen, zoals afwijkingen in de gegevens of onverwachte/onverklaarbare resultaten.
    - □ Na elke oplevering is beschreven of er veranderingen zijn in de gegevens en of er afwijkende resultaten zijn.
    - □ Waarnemingen zoals afwijkingen in gegevens of onverwachte/onverklaarbare resultaten zijn genoteerd.

- Gebruik eenvoudige methoden boven complexe methoden daar waar mogelijk. Dit komt ten goede aan de uitlegbaarheid, auditeerbaarheid en beperking van risico's.
    - □ De meest simpele methode die toch het doel van het algoritme bereikt is gebruikt.
    - □ Onderbouw de keuze waarom voor kunstmatige intelligentie technieken is gekozen, bijv. omdat dit tot veel betere resultaten leidt dan een simpele methode.
    - □ Zonder de complexe methode behaalt het algoritme niet het doel waarvoor het ontwikkeld is.

- Verifieer zowel voor statistische analysemethoden als complexere methoden de uitkomsten gebaseerd op de specifieke input. Door beide methoden te gebruiken is het mogelijk om te herleiden of een meer complexe analyse tot betere uitkomsten leidt. Indien dit niet het geval is, heeft het de voorkeur om een eenvoudigere statistische methode in te zetten.
    - □ Onderdeel van de modelontwikkeling is validatie waarin een aantal verschillende methoden zijn gebruikt en geverifieerd.
    - □ Ontwikkel bijv. een baseline model dat intrinsiek uitlegbaar is m.b.v. lineaire regressie/logistische regressie/decision tree.
    - □ Vergelijk vervolgens de complexere modellen met het baseline model.
    - □ Gebruik methodes om uitlegbaarheid van complexe modellen te vergroten

- Lever een gedetailleerde omschrijving van het model en de werking ervan, samen met een controleerbare validatie dat de code overeenkomt met de specificatie.
    - □ In de technische documentatie is de werking van het model gedetailleerd beschreven.
    - □ Beschrijf de werking van het model tot in detail.
    - □ Tevens is hier van belang het documenteren/annoteren in de code wat een routine doet of moet doen om een goede code-review uit te kunnen laten voeren door bijv. een andere analist.

- Zorg voor reproduceerbaarheid.
    - □ De resultaten van het model zijn reproduceerbaar.
    - □ Codeer alle stappen in de analyse. Zorg er voor dat er geen handmatige acties nodig zijn.
    - □ Zorg voor versiebeheer op de code.
    - □ Bewaar geanonimiseerde trainingsdata.
    - □ Documenteer alle aannames, parameterkeuzes en andere beslissingen, en leg ook de onderbouwing vast.

- Zorg dat er bij algoritmen die in besluitvorming uitmonden dan wel anderszins een aanmerkelijke impact op burgers, bedrijven of samenleving, een vorm van betekenisvolle menselijke tussenkomst is georganiseerd en leg de invulling van deze menselijke tussenkomst vast, zodat die tussenkomst in praktijk ook daadwerkelijk betekenisvol is.[35]

[35] Dat in deze situaties sprake moet zijn van menselijke tussenkomst volgt uit artikel 22 AVG (voor wat betreft data-analyses waarin persoonsgegevens worden gebruikt). Er dient dan wel sprake te zijn van een situatie als bedoeld in artikel 22 AVG. Dit wil zeggen:
a. Geautomatiseerde individuele besluitvorming met profilering (eerste lid);
b. Geautomatiseerde individuele besluitvorming die noodzakelijk is voor de totstandkoming of uitvoering van een overeenkomst (tweede lid, onder a);

5. Verantwoording

Om verantwoording af te kunnen leggen moet een dossier aangelegd worden met alle relevante informatie betreffende het ontwikkelen en operationaliseren van het model. Overheden zijn verantwoordelijk voor de ontwikkeling en inzet van hun algoritmen en dienen daarover dan ook verantwoording af te leggen.

Formeel is de betreffende minister verantwoordelijk voor de taakuitoefening door een onderdeel van de rijksdienst. In de praktijk zal deze bevoegdheid zijn gemandateerd, bijvoorbeeld aan een directeur-generaal of een directeur van een dienstonderdeel. De gemandateerde functionaris is dan verantwoordelijk voor de inzet van algoritmen en de toepassing daarbij van de richtlijnen en andere relevante instrumenten en normen. Bij decentrale overheden of ZBO's met publiekrechtelijke rechtspersoonlijkheid geldt eveneens dat een verantwoordelijkheid van de bestuurder gemandateerd kan zijn naar functionarissen van een dienstonderdeel.

- Wanneer meerdere organisaties betrokken zijn bij de ontwikkeling van algoritmische systemen, wordt geregeld dat één organisatie het voortouw en regie heeft bij de ontwikkeling van het systeem en als zodanig primair voor de toepassing van de richtlijnen en de uitvoering van andere relevante instrumenten, zoals DPIA en AI mensenrechten impact assessment.[36]

- Bij complexe algoritmische systemen dient binnen de eigen organisatie de regie en bijbehorende verantwoordelijkheid voor het ontwerp- en ontwikkelproces duidelijk te worden belegd (bij een team of persoon, bijv. projectleider). Verantwoording wordt afgelegd in de hiërarchische lijn richting de verantwoordelijk minister/ bestuurder. Daarbij wordt geborgd dat ook na implementatie van het systeem (wanneer het systeem een lijn-activiteit is geworden), de regie en verantwoordelijkheid helder belegd zijn en in een governance structuur zijn vastgelegd.

- De verantwoording gebeurt volgens het principe van comply or explain. Dit wil zeggen dat als uitgangspunt geldt dat organisaties de richtlijn moet volgen, of nadrukkelijk uitleggen waarom zij ervan afwijken. Dit geldt in eerste instantie voor algoritmen die nieuw worden ontwikkeld, maar ook voor algoritmen die reeds werkzaam zijn.

- Gelet daarop wordt de toepassing van de richtlijnen bij voorkeur verankerd in het kader van een P&C-cyclus, waarbij jaarlijks door organisaties wordt gerapporteerd over situaties waarin besloten is om af te wijken van de richtlijnen.

- Zorg dat de voor verantwoording noodzakelijke informatie (zoals documentatie, broncode of data) beheerd wordt conform de eisen die de Archiefwet stelt.

6. Validatie

Overheden moeten gebruik maken van strikte methoden om hun modellen te valideren en deze methoden en resultaten documenteren. Overheden worden aangemoedigd om deze resultaten openbaar te maken. In het bijzonder moeten ze routinematig testen uitvoeren om te beoordelen en bepalen of het model het beoogde doel en functionaliteit bereikt en geen bijkomende schade oplevert.

- □ De overheidsinstantie gebruikt strikte methoden om het model te valideren.
- □ Modellen met een grote potentiële impact worden gevalideerd door een onafhankelijke partij.
- □ De overheidsinstantie documenteert de methode en de resultaten.
- □ De overheid test en beoordeelt routinematig of het model het beoogde doel bereikt en geen bijkomende schade oplevert.
- □ Het model wordt continu gemonitord.
- □ Uitgangspunt is dat de overheidsinstantie over deze testresultaten publiceert.

---

c. Geautomatiseerde individuele besluitvorming die berust op de uitdrukkelijke toestemming van betrokkene (tweede lid, onder b).

Ook bij voorspellende algoritmen (risicotaxatie-modellen) die niet in besluitvorming uitmonden (maar in een risicomelding zoals bij het SyRi systeem) en een aanmerkelijke impact hebben op burgers, verdient het de voorkeur om (betekenisvolle) menselijke tussenkomst te organiseren.

[36] Deze organisatorische afspraken doen niet af aan de eigen cq. gezamenlijke verantwoordelijkheid van de andere betrokken ministers/organisaties. Zie artikel 26 AVG en 20 Richtlijn.

7. Toetsbaarheid

Waar het bij 'uitlegbaarheid' gaat om het in begrijpelijke taal beschrijven van de uitkomsten van de analyse, ziet toetsbaarheid erop dat de uitkomsten daadwerkelijk getoetst kunnen worden. Data-analyse dient zo te worden ingericht dat de methode van data-analyse, de gehanteerde algoritmen, datasets en de feitelijke verwerkingen daadwerkelijk kunnen worden getoetst. In het bijzonder wanneer data-analyse wordt toegepast ten behoeve van besluitvorming. In dat geval dient het algoritme, te beschikken over een toereikend niveau van transparantie, verifieerbaarheid en toetsbaarheid. Toetsbaarheid richt zich primair op toetsing door de toezichthouder en de rechter. Vanuit de Audit Dienst Rijk (ADR) is een Normenkader auditing algoritme opgesteld en is de Algemene Rekenkamer (AR) voornemens een toetsingskader voor algoritmen op te stellen. De bestuursrechter stelt op grond van het bestuursrecht eisen aan de inzichtelijkheid, controleerbaarheid en toegankelijkheid in geval van geautomatiseerde besluitvorming door bestuursorganen. Voor het toetsingskader voor de beoordeling van geautomatiseerde besluitvormingsprocessen m.b.v. algoritmen, zie de uitspraak van de Afdeling bestuursrecht van de Raad van State van 17 mei 2017 en voornoemd ongevraagd advies van de Raad van State van 31 augustus 2018. [37]

[37] ECLI:NL:RVS:2017:1259, i.h.b. rechtsoverwegingen 14.3 en 14.4, https://www.recht.nl/rechtspraak/uitspraak/?ecli=ECLI:NL:RVS:2017:1259.
Ongevraagd advies van de Raad van State over de effecten van digitalisering voor de rechtstatelijke verhoudingen', Kamerstukken II, 2017/18, 26643, nr. 557

Richtlijn inzake publieksvoorlichting over data-analyses

Deze richtlijn heeft vooral betekenis in het geval dat bij het uitvoeren van data-analyses persoonsgegevens worden verwerkt. Gelet daarop wordt ook aandacht besteed aan relevante voorschriften uit de AVG.

Strekking richtlijn
Artikel 5, eerste lid, onder a, AVG bepaalt dat persoonsgegevens worden verwerkt op een wijze die ten aanzien van betrokkenen transparant is. Dit beginsel wordt uitgewerkt in verschillende informatieverplichtingen. Een algemeen kader daarvoor is te vinden in artikel 12 AVG. Waar deze verplichtingen zich op de 'betrokkenen' richten, ziet deze richtlijn tevens op de informatievoorziening aan 'het publiek' en kunnen in die zin worden gezien als een aanvulling op het transparantiebeginsel uit de AVG, toegesneden op de specifieke kenmerken van algoritmische data-analyses.

Het gaat in deze richtlijn louter om algemene informatie, bestemd voor het publiek. Verplichtingen uit de AVG om een betrokken burger individueel over verwerking van hem betreffende persoonsgegevens te informeren[38], blijven hier dus buiten beschouwing.

De gewenste mate van publieksvoorlichting zal altijd gebonden zijn aan de context waarin de data-analyses plaatsvinden: een overheid als dienstverlener zal doorgaans meer transparantie kunnen betrachten dan een overheid die als toezichthouder of opsporingsinstantie optreedt.

Algemene voorwaarden aan transparantie
De algemene informatie die over data-analyses aan het publiek en betrokkenen wordt verstrekt, dient, waar het de verwerking van persoonsgegevens betreft, aan een aantal voorwaarden te voldoen.

Waarover transparant?
Als een overheidsdienst data analyses verricht, dient deze dienst op haar website het publiek te informeren over:
- dat zij data-analyses uitvoert;
- waarom zij data-analyses uitvoert (wat het doel ervan is en wat met de resultaten daarvan wordt gedaan);
- waarom het gebruik van data analyses proportioneel is, en er geen betere alternatieven waren om het doel te bereiken;
- wat de eventuele consequenties van de analyse voor betrokken burgers zijn, en hoe rekening is gehouden met eventuele impacts ervan op grondrechten;
- eventuele toepassing van *machine learning* en de uitleg daarvan;
- wat de wettelijke grondslag voor het uitvoeren van deze analyses is,
- welke databronnen van welke organisaties daarvoor worden gebruikt, en wat de kwaliteit daarvan is;
- welke persoon binnen de overheidsorganisatie verantwoordelijk voor de analyse is;
- wat de rol van eventuele derden bij deze analyses is;
- met welke organisaties (publiek of privaat), indien dit aan de orde is, brondata en/of resultaten van de analyses worden gedeeld en op basis van welke grondslag;
- welke kwaliteitsborging er plaatsvindt (welke risico's worden onderkend en welke maatregelen daartegen worden genomen en op welke wijze toetsing plaatsvindt);
- hoe er tussen analyse en een eventueel besluit menselijke tussenkomst plaatsvindt, en welke toetsingskaders er zijn.

Zij moet in de eerste plaats beknopt en transparant, begrijpelijk en gemakkelijk toegankelijk zijn. "Beknopt en transparant" wil in dit verband zeggen dat de informatie efficiënt en bondig moet worden gepresenteerd. Als een overheidsdienst de informatie in een *privacy statement* op haar website opneemt, kan dit betekenen dat de informatie daarin "gelaagd" wordt opgenomen, waarbij men snel kan navigeren naar relevante passages, zonder door het gehele privacy statement te hoeven scrollen. "Begrijpelijk" impliceert dat een gemiddelde vertegenwoordiger van het beoogde publiek de informatie moet kunnen begrijpen. Het kan in dit verband nuttig zijn om af en toe bij het actuele publiek te checken of dit het geval is, bijvoorbeeld door middel van een gebruikerspanel. "Gemakkelijk toegankelijk" betekent dat men weinig moeite hoeft te doen om toegang tot de informatie te krijgen. Opneming daarvan in een *privacy statement* op de eigen website met een duidelijke link op de *homepage* kan

[38] Zie de artikelen 13 en 14 AVG.

daaraan bijdragen. De informatie moet ook in duidelijke en eenvoudige taal worden verschaft, vooral als de informatie voor kinderen is bestemd. Zij moet geen ruimte voor verschillende interpretaties geven en in ieder geval duidelijkheid verschaffen over het doel en de wettelijke grondslag van de analyses.[39]

Inzichtelijkheid algoritmen
Een overheidsdienst kan ervoor kiezen de algoritmen die zij voor haar data-analyses gebruikt, inzichtelijk te maken voor het publiek, maar kan daarmee niet volstaan. Een burger zal deze algoritmen immers doorgaans niet of nauwelijks begrijpen. Informatie over de toepassing van algoritmen en over het proces van bijvoorbeeld kwaliteitsbewaking kunnen de werking van het algoritme inzichtelijker maken. Dit laat onverlet dat een overheidsorganisatie zelf steeds inzicht moet hebben over hoe een algoritme werkt en hierover telkens wanneer dat nodig is – bijvoorbeeld in het kader van individuele besluitvorming – verantwoording moet kunnen afleggen.

Cristal box
Om de transparantie rond data-analyses te vergroten zou een overheidsdienst een zgn. C*ristal box* kunnen ontwikkelen waarmee je inzicht in de analyse kunt krijgen en daarop controle kunt uitoefenen, zonder dat je van de technische details op de hoogte hoeft te zijn. Zo'n *Cristal box* geeft inzicht in het analyseproces, de gebruikte algoritmen en de gebruikte datasets. Zo kan men beter achterhalen welke beslissingen in het analyseproces zijn genomen en welke variabelen zijn gebruikt.

Hoe groter de gevolgen, des te belangrijker de transparantie
Naarmate de gevolgen van Data-analyses voor burgers groter zijn, is transparantie belangrijker. Dat geldt zeker in gevallen waarin dergelijke analyses tot conclusies of besluiten (kunnen) leiden.

Getrapte transparantie
De gehanteerde transparantie kan getrapte zijn. Dit betekent dat, als een overheidsdienst over haar data-analyses wel in algemene termen transparantie betracht maar bepaalde, meer gedetailleerde aspecten met het oog op bijvoorbeeld het belang van de opsporing niet openbaar maakt, zij een dergelijke handelswijze tenminste compenseert met voldoende intern en extern toezicht op die aspecten.

Voorkom *gaming the system*
Bij het betrachten van transparantie zal rekening moeten worden gehouden met het risico van *gaming the system*: wanneer (veel) inzicht in methoden en data wordt gegeven kunnen kwaadwillenden hier misbruik van maken. Zo kan het wenselijk zijn dat een overheidsdienst in haar privacy statement wel informatie verschaft over variabelen die zij in haar analyses hanteert, maar niet over drempelwaarden. Dit laat onverlet dat overheidsdiensten bij verwerking van persoonsgegevens in de data-analyse op grond van de AVG een betrokken burger individueel over verwerking van hem betreffende persoonsgegevens moet informeren. Maar ook dat zij dan op grond van sectorspecifieke wetgeving deze verplichting mogelijk categorisch buiten toepassing kunnen laten dan wel op grond van artikel 41 van de Uitvoeringswet AVG in individuele zaken buiten toepassing kunnen laten.[40]

Proeftuinen en experimenteerruimten
Als een overheidsdienst in een zgn. proeftuin met data-analyses wil gaan experimenteren, is het wenselijk in dat kader ook duidelijkheid te verkrijgen over de wijze waarop men transparantie kan en wil betrachten. Bij twijfel hierover is het wenselijk daarover de discussie aan te gaan en naar buiten te treden.

Transparantie en toetsbaarheid
Transparantie van algoritmen voor het grote publiek en de mogelijkheid tot toetsing van een algoritme dienen niet met elkaar te worden verward. Ook al zou het weinig zin hebben om transparant te zijn over de gehanteerde algoritmen zelf, zij dienen wel toetsbaar te worden gemaakt. Transparantie en toetsbaarheid hebben in zoverre weer wel met elkaar te maken

[39] Zie nader artikel 12, eerste lid, AVG en de Guidelines on transparency under Regulation 2016/679 (AVG) van de Article 29 Data protection working party, paragraaf 6-18.
[40] Bij een beroep op artikel 41 UAVG zal dan moeten worden aangetoond dat dit noodzakelijk of evenredig is ter waarborging van een aantal in dat artikel genoemde belangen. Daartoe behoren, kort gezegd, onder meer de nationale veiligheid, de landsverdediging, de openbare veiligheid, de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten en toezicht of inspectie.

dat, indien een data-analyse door een deskundige “onafhankelijke” partij is getoetst en in orde bevonden, het wenselijk is het publiek daarvan in kennis te stellen, zodat een burger zelf niet alles tot in detail hoeft te begrijpen.