Auditdienst Rijk
Ministerie van Financiën

Onderzoeksrapport beheersing algoritmes binnen het Rijk

Definitief

Colofon

Titel	Onderzoeksrapport beheersing algoritmes binnen het Rijk
Uitgebracht aan	CIO-Rijk
Datum	8 juni 2022
Kenmerk	2022-0000168432

Inlichtingen
Auditdienst Rijk
070-342 7700

Inhoud

1 Inleiding

1.1 Aanleiding

De ministeries van Binnenlandse Zaken en Koninkrijksrelaties (BZK) en Justitie en Veiligheid (JenV) hebben in 2019 een onderzoek laten uitvoeren naar het toezicht op algoritmes door de overheid. Uit dit onderzoek bleek dat er in de praktijk nog geen sprake is van structureel onderzoek door toezichthoudende organisaties naar het gebruik van algoritmes door de overheid. Een aanbeveling was om overheids-organisaties periodieke audits uit te laten voeren. De minister van BZK en de minister voor Rechtsbescherming hebben in dit kader in hun reactie aangegeven dat "met de ADR zal worden verkend welke rol zij kan spelen bij het op structurele basis uitvoeren van audits op het algoritmegebruik door de Rijksoverheid". Hierna zijn beide ministeries en de ADR samen opgetrokken om een eerste invulling hieraan te geven met een Rijksbreed onderzoek. Ook in het vorig jaar verschenen rapport van de Algemene Rekenkamer (AR) werd aandacht gevraagd voor de kwaliteit van het testen van de algoritmes en de continue monitoring door het ministerie.

De ADR heeft een normenkader ontwikkeld om de (beheersing van de) kwaliteit en verantwoorde inzet van een algoritme door een departement en daaronder vallende organisaties te onderzoeken. Het gebruik van algoritmes biedt kansen voor departementen om hun werkprocessen en/of dienstverlening te verbeteren, maar brengt ook risico's met zich mee als de algoritmes niet goed beheerst worden. De CIO Rijk deelt dit beeld en heeft als opdrachtgever (namens het CIO-Beraad) de ADR gevraagd om Rijksbreed onderzoek te doen naar de huidige wijze van beheersing van algoritmes en waar nodig aanbevelingen te geven voor aanvullende maatregelen bezien vanuit zijn stelselverantwoordelijkheid.

1.2 Opdracht

Doelstelling
De doelstelling van dit onderzoek naar de beheersing van algoritmes binnen het Rijk is "inzicht geven in de kwaliteit van de beheersing van de geselecteerde algoritmes door onderzoek te doen naar de wijze waarop de organisaties de kwaliteit en verantwoorde inzet van hun algoritmes borgen. Op basis van deze inzichten en eventuele aanbevelingen kan de opdrachtgever verbeteringen doorvoeren op stelselniveau".

Onderzoeksvragen
De doelstelling leidt tot de volgende centrale onderzoeksvraag:

Op welke wijze is vormgegeven aan de beheersing van het algoritme om de kwaliteit en verantwoorde inzet ervan te borgen?

In dit onderzoek worden de volgende deelvragen beantwoord:

1. Op welke wijze is de governance rond het algoritme ingericht?
2. Op welke wijze is invulling gegeven aan ethische principes?
3. Op welke wijze is de privacy geborgd?
4. Op welke wijze is de kwaliteit van de General IT Controls (GITC) geborgd?
5. Op welke wijze is de kwaliteit van het model geborgd?
6. Op welke wijze is de kwaliteit van de input- en outputdata geborgd?
7. Welke handelingsperspectieven kunnen worden gegeven om de kwaliteit van de beheersing van algoritmes te verbeteren?

Achtergrond onderzochte algoritmes

Het onderzoek heeft zich gericht op vijf algoritmes bij verschillende departementen en uitvoeringsorganisaties. Deze algoritmes bevonden zich in verschillende fases: ontwerp en ontwikkeling (incl. pilots), toepassing (in productie) en beheer. Tevens was één algoritme bij een externe partij in ontwikkeling en in beheer. Het type algoritme varieerde van relatief eenvoudigere beslisbomen tot machine learning en neurale netwerken die over het algemeen complexer zijn in techniek en uitlegbaarheid. Alle algoritmes waren ondersteunend aan beslissingen die medewerkers moesten nemen. De impact die ze hadden op de burger en maatschappij was beperkt. Binnen deze context hebben wij ons onderzoek uitgevoerd en per onderzocht algoritme en organisatie een deelrapportage in concept uitgebracht[1]. Op basis van deze rapportages en onderliggende analyses hebben wij de rode draden gevonden zoals weergeven in dit rapport.

1.3 Leeswijzer

Hoofdstuk 2 betreft de managementsamenvatting met de bevindingen en aanbevelingen op hoofdlijnen. In hoofdstuk 3 gaan we in op de wijze waarop de governance rond algoritmes is ingericht en op de GITC (deelvraag 1 en 4). In hoofdstuk 4 gaan we in op de wijze waarop invulling is gegeven aan ethische principes (deelvraag 2) en privacy (deelvraag 3). Hoofdstuk 5 gaat over de kwaliteit van het model (deelvraag 5) en van de input- en outputdata (deelvraag 6). Het handelingsperspectief in de vorm van aanbevelingen (deelvraag 7) is bij de betreffende hoofdstukken opgenomen. Tot slot is in hoofdstuk 6 de verantwoording van het onderzoek opgenomen.

[1] De deelrapportages wachten nog op een formele status “uitgebracht”, derhalve zijn deze nog in concept.

2 Managementsamenvatting

Wij hebben bij vijf organisaties onderzoek gedaan naar de wijze waarop vorm is gegeven aan de beheersing van een algoritme. Op basis van de vijf deelrapportages is dit onderzoeksrapport opgesteld met rode draden. Dit rapport geeft inzichten de beheersing van algoritmes en geeft aanbevelingen op basis van waarvan de opdrachtgever verbeteringen kan doorvoeren op stelselniveau.

Uit de deelonderzoeken komt het beeld naar voren dat de wijze van beheersing van algoritmes verschilt en onder meer afhangt van de fase waarin het algoritme zich in bevindt (van ontwerp tot beheer), de complexiteit, autonomie en impact van het algoritme en of deze intern of extern ontwikkeld en beheerd is.

Over het algemeen zien we dat de techniek achter het algoritme doordacht is. De organisaties kunnen overwegingen en keuzes ten aanzien van het model en de privacy en ethische aspecten in de meeste gevallen onderbouwen, maar deze komen niet altijd voort uit een gestructureerde risicoanalyse en zijn vaak beperkt vastgelegd. Daarnaast zijn de criteria waaraan het algoritme moet voldoen meestal niet vooraf bepaald. Dit maakt het lastig om achteraf als organisatie te beoordelen hoe het algoritme functioneert en daarover verantwoording af te leggen. Met andere woorden, de plan-do-check-act (PDCA) cyclus van algoritmes is vaak nog niet sluitend. We bevelen aan om een handreiking op te stellen over hoe organisaties hun PDCA-cyclus rondom algoritmes in samenhang met de geformuleerde doelstelling(en) kunnen vormgeven. Dit moet bijdragen aan meer grip op algoritmes.

Hieronder geven we de bevindingen en aanbevelingen op hoofdlijnen. Deze zijn in de hoofdstukken 3 t/m 5 nader toegelicht.

1. Risico's, afwegingen en getroffen maatregelen beperkt inzichtelijk

Het is doorgaans helder hoe de algoritmes bijdragen aan de taakuitvoering van de organisaties. De algoritmes zijn ontstaan vanuit innovatieve ideeën om primaire werkprocessen te verbeteren. Voorafgaand aan de ontwikkeling zijn het doel, de wensen en eisen vanuit de business (opdrachtgever/eindgebruikers) globaal bekend. Een daaropvolgende gestructureerde risicoanalyse met de business en ontwikkelaars mist veelal. Uit documenten en interviews is op te maken dat aandacht is besteed aan diverse risico's omtrent het algoritme, maar de afwegingen en maatregelen om deze risico's te mitigeren, zijn in beperkte mate uitgewerkt en vastgelegd. Het is daardoor niet geheel inzichtelijk of alle risico's bekend en beheerst zijn.

Aanbeveling
Geef in een handreiking aan wanneer en hoe risicoanalyses rond het algoritme kunnen worden uitgevoerd, o.a. volgend op de wensen en eisen uit de business.

2. Analyse van de privacyrisico's en de impact kunnen scherper

De organisaties die bij de data-analyse gebruik maken van persoonsgegevens zijn zich hiervan bewust en hebben oog voor privacy aspecten. Een *data protection impact assessment* (DPIA) is echter niet altijd (volledig) uitgevoerd indien het algoritme persoonsgegevens verwerkt. Daar waar een DPIA deels of volledig is uitgevoerd, is een passende keuze gemaakt om het algoritme als onderdeel van de gehele verwerking mee te nemen. Daarbij merken we op dat de samenhang tussen de uitwerking van risico's en maatregelen nog beperkt is. Het niet tijdig of volledig

uitvoeren van een DPIA kan ertoe leiden dat een organisatie niet alle privacyrisico's in beeld heeft en daardoor het risico loopt niet te voldoen aan wet- en regelgeving.

Vanuit ethisch oogpunt is de impact van het algoritme op burgers, bedrijven en maatschappij over het algemeen helder voor de betreffende organisatie, al verschilt de diepgang waarmee dit is geanalyseerd en vastgelegd. De meeste algoritmes zijn relatief eenvoudig, wat de uitlegbaarheid van het algoritme vergroot.

Aanbeveling
Geef in een handreiking aan wanneer en door wie risico's t.a.v. privacy en ethiek bij algoritmes geanalyseerd en beheerst kunnen worden.

3. Veel aandacht voor kwaliteit van model en data, maar een eenduidige methodiek om dit te beoordelen mist

De organisaties zien op verschillende wijze toe op de kwaliteit van de data en het model. Daarbij heeft de kans op bias en discriminatie door het algoritme aandacht van de organisaties gekregen gedurende het ontwikkelproces. Een aandachtspunt betreft de monitoring van mogelijke discriminatie bij het gebruik van het algoritme. De organisatie moeten blijven toezien of het algoritme nog nauwkeurig en robuust genoeg is. Het blijkt dat een eenduidige methode om te toetsen of een gebruikte dataset (nog) representatief genoeg is, nog niet aanwezig is. Dit geldt ook voor de controle of de datasets juist, volledig en consistent zijn. Wanneer de data en/of het algoritme (model) niet goed genoeg meer zijn, lopen organisaties het risico dat het algoritme discrimineert en/of het doel niet bereikt.

Aanbeveling
Overweeg een gerichte aanpak voor een eenduidige methode waarmee de kwaliteit van het model (nauwkeurigheid en robuustheid) en de data (representativiteit en juistheid/volledigheid/consistentie)onderbouwd kan worden.

4. Controle op en verantwoording over algoritmes niet helder

Het is belangrijk dat organisaties verantwoording afleggen over de toegepaste algoritmes. Daarvoor moeten ze gedurende het gebruik nagaan hoe goed een algoritme functioneert aan de hand van vooropgestelde criteria, zodat eventueel bijgestuurd kan worden. Wij constateren dat de onderzochte organisaties wel meten hoe een algoritme presteert, maar dat zij dit veelal niet afzetten tegen vooraf opgestelde criteria. De acceptatie- en prestatiecriteria zijn vaak niet bepaald en/of beschreven. Ook ontbreekt vaak een controle voorafgaand aan de implementatie of aan de relevante wet- en regelgeving en (interne) richtlijnen is voldaan, evenals of de gemaakte keuzes overeenkomen met de wensen en eisen van de business.

Afspraken over de wijze waarop organisaties het algoritme monitoren en evalueren zijn nog beperkt gemaakt. Dit laatste geldt ook voor algoritmes die extern zijn ontwikkeld en beheerd. Daarbij hebben wij bevonden dat geen eenduidigheid bestaat wie voor en na implementatie de verantwoordelijkheid heeft voor het algoritme. Hierdoor bestaat het risico dat (bij)sturing en verantwoording afleggen niet correct en tijdig gebeurt. Verantwoordings- en rapportagelijnen tussen de betrokken functionarissen moeten veelal nog vastgesteld worden.

Aanbeveling
Geef in de handreiking aan:
- Wanneer en door wie de controle wordt uitgevoerd of het algoritme voldoet aan relevante wet- en regelgeving en (interne) richtlijnen.
- Wanneer en aan de hand van welke vooropgestelde criteria (o.a. acceptatie- en prestatiecriteria) het algoritme gemonitord en geëvalueerd kan worden.
- Door wie en hoe verantwoording over de inzet en het gebruik van het algoritme afgelegd dient te worden.

3 Governance

In dit hoofdstuk beschrijven we de wijze waarop de governance rond het algoritme is ingericht (deelvraag 1) en de GITC (deelvraag 4). De onderstaande paragrafen geven inzicht in de wijze van sturing, beheersing en verantwoording.

3.1 Beleid voor inzet algoritmes is in ontwikkeling

Het onderzoek laat een eenduidig beeld zien dat een beleid ten aanzien van algoritmes nog niet aanwezig is. Hierin kan de wijze waarop en het doel waarmee organisaties algoritmes inzetten worden beschreven, met als doel de kwaliteit en verantwoorde inzet van het algoritme te optimaliseren. Een aantal organisaties geeft aan dat een dergelijk beleid momenteel in ontwikkeling is. Wij merken op dat beleid volgend is aan de ontwikkelingen. In beginsel lag de focus vooral op het innoveren: het onderzoeken van de mogelijkheden om algoritmes in bestaande processen toe te passen. Nu de organisaties verder zijn en in grotere mate gebruik (gaan) maken van algoritmes, lijkt de behoefte aan beleid toe te nemen.

Aangegeven is dat bij de ontwikkeling en het gebruik van de algoritmes soms wel andere gerelateerde beleidskaders worden geraadpleegd, zoals op het gebied van privacy of informatiebeveiliging. In deze beleidskaders wordt zelden expliciet verwezen naar algoritmes, maar veel aspecten die in deze kaders terugkomen zijn ook relevant bij het gebruik van algoritmes. Desalniettemin zijn afspraken over de manier waarop en voor welk doel algoritmes worden ingezet nog nodig.

3.2 Doelstelling is veelal helder, de afweging over de doelmatigheid is niet altijd aanwezig

Het doel van de algoritmes en de bijdrage aan het proces zijn helder en veelal beschreven. De onderzochte algoritmes worden ingezet in werkprocessen waarbij data voorheen veelal handmatig geanalyseerd werd. De algoritmes dragen er mede aan bij om de data sneller te verwerken en/of om slimmere keuzes te maken. Een volgende stap is om de doelen meer SMART te maken. Dit maakt betere sturing op en verantwoording over de inzet van het algoritme mogelijk.

Daarnaast zien we dat vaak geen bewuste afweging is gemaakt of het algoritme het juiste middel is om de taakuitvoering op doelmatige wijze te realiseren. De keuze voor het in te zetten algoritme, waarbij de voorkeur uitgaat naar een eenvoudig algoritme zoals een beslisboom boven een complexer algoritme, is vaak wel beargumenteerd. We zien met name dat niet in alle gevallen een goede afweging is gemaakt over de kosten en baten van de inzet van het algoritme.

Het is van belang om naast het doel en het type algoritme als organisatie ook te bepalen welke gegevens het algoritme gaat verwerken en wat de gehanteerde (hyper)parameters, acceptatiecriteria en prestatiecriteria zijn. Organisaties blijken de bovengenoemde aspecten veelal niet vast te stellen en vast te leggen. Daarmee ontbreekt een essentiële basis om het algoritme te evalueren. Wanneer vaststelling ontbreekt, kunnen mogelijk onderlinge interpretatieverschillen ontstaan wat kan leiden tot onbedoelde nadelige effecten en andere gebreken.

3.3 Risico's, maatregelen, overwegingen niet altijd inzichtelijk

Organisaties dienen risicoanalyses uit te voeren voorafgaand aan en gedurende de (door)ontwikkeling van een algoritme. Dit is onder meer van belang om te zorgen dat een algoritme bijdraagt aan het beoogde doel, voldoet aan compliance vereisten

en uitlegbaar en verantwoordbaar is. Over het algemeen blijkt uit de interviews en documenten dat de organisaties aandacht hebben voor risico's. Het gaat dan om zowel operationele risico's (gericht op realisatie van een project en implicaties voor de uitvoering) als ook de technische risico's (gerelateerd aan de kwaliteit van het algoritme) en risico's met het oog op ethische en privacy principes. De meer technische risico's in relatie tot de impact van het algoritme verdienen bij de risicoanalyse veelal nog aandacht.

De onderzochte organisaties hebben diverse activiteiten uitgevoerd om inzicht te krijgen in deze risico's, al verschilt de diepgang waarmee dit is gedaan wel. Zo hebben enkele organisaties een *data protection impact assessment* (DPIA) opgesteld om privacyrisico's te inventariseren en beoordelen (zie paragraaf 4.1). Daarnaast zijn andere beschikbare richtlijnen en toetsingskaders geraadpleegd. Voorbeelden zijn de richtlijnen van Justitie en Veiligheid, het Artificial Intelligence Impact Assessment (AIIA) van het ECP en de kaders van de Algemene Rekenkamer en ADR. Enkele organisaties hebben zelf een checklist opgesteld aan de hand van deze kaders.

Wat opvalt is dat de geïnventariseerde risico's beperkt of versnipperd zijn vastgelegd. Daarnaast is niet altijd duidelijk hoe de organisaties zijn omgegaan met de adviezen en aandachtspunten die uit de genoemde activiteiten en analyses naar voren zijn gekomen. Dit maakt het totaalbeeld van risico's en getroffen maatregelen niet eenvoudig raadpleegbaar. Dit kan de monitoring ervan bemoeilijken en vormt ook een risico voor de overdracht van kennis. Hetzelfde geldt in die zin voor de vastlegging van gemaakte keuzes en afwegingen bij de totstandkoming van het algoritme.

Tevens is in de meeste gevallen nog niet geheel vastgesteld dat het algoritme voldoet aan relevante wet- en regelgeving en (interne) richtlijnen, zowel op het gebied van privacy en ethiek als op het specifieke beleidsdomein. Daarbij dient opgemerkt te worden dat een aantal algoritmes nog niet in gebruik is en organisaties aangeven voornemens zijn om voor de implementatie voorgenoemde vast te stellen. Door dit later te doen, bestaat het risico dat niet voldaan wordt aan wet- en regelgeving en eerder gemaakte keuzes tijdens de ontwikkeling herzien moeten worden of momenteel niet juist en rechtmatig worden ingezet.

Aanbeveling

Geef in een handreiking aan wanneer en hoe risicoanalyses (bijvoorbeeld middels een bepaalde methodiek of format) rond het algoritme kunnen worden uitgevoerd en vastgelegd. Hierbij moet rekening gehouden worden met het doel, de wensen en eisen van de business en de verschillende (ethische) aspecten van een algoritme. Daarnaast is het goed om aan te geven wanneer en hoe controle wordt uitgevoerd of het algoritme voldoet aan relevante wet- en regelgeving en (interne) richtlijnen. Denk daarbij aan de reeds beschikbare instrumenten.

3.4 Monitoring en evaluatie van het algoritme moet nog worden ingevuld

Afspraken over de wijze waarop het algoritme wordt gemonitord en geëvalueerd zijn nog beperkt gemaakt. Twee organisaties, waarvan het algoritme nog niet in productie is, hebben hier enige invulling aan gegeven. Zo is eenmaal een advies voor monitoring en evaluatie (uitgebreid) uitgewerkt, maar zijn nog niet opgepakt en vastgesteld. De organisaties die het algoritme reeds toepassen hebben nog weinig duidelijke afspraken intern en/of met de leverancier gemaakt over de wijze van evalueren. Hierdoor is weinig tot geen inzicht in de doeltreffendheid, doelmatigheid en eventuele ongewenste nadelige effecten.

De monitoring en evaluatie worden bemoeilijkt doordat vooraf veelal geen acceptatiecriteria en prestatiecriteria zijn bepaald en/of beschreven. Dit zijn criteria die aangeven hoe goed het algoritme moet presteren gedurende het gebruik. Deze

criteria zouden moeten volgen uit een brede risicoanalyse en -afweging voorafgaand aan het gebruik van het algoritme, in overleg met de opdrachtgever. Deze criteria maken onder meer duidelijk hoe nauwkeurig het algoritme moet zijn en welk percentage fouten acceptabel is voor de opdrachtgever wat weer essentieel is voor de evaluatie.

Aanbeveling

Geef in een handreiking aan wanneer en aan de hand van welke vooropgestelde criteria (o.a. acceptatie- en prestatiecriteria) het algoritme gemonitord en geëvalueerd kan worden. Deze moeten volgen uit een risicoanalyse en -afweging waarbij de opdrachtgever en opdrachtnemer betrokken zijn.

3.5 Taken, bevoegdheden en verantwoordelijkheden zijn doorgaans helder bij ontwikkeling, maar nog niet altijd bij controle en verantwoording

Bij de ontwikkeling van de algoritmes zien we dat de taken, bevoegdheden en verantwoordelijk over het algemeen belegd zijn. Verschillende relevante afdelingen vanuit de opdrachtgever en -nemer (zoals ontwerpers, ontwikkelaars, beheerders, eindgebruikers) zijn betrokken bij de ontwikkeling van het algoritme. Hoewel hun taken, bevoegdheden en verantwoordelijkheden veelal niet zijn beschreven, zijn deze volgens geïnterviewden doorgaans wel helder. De betrokken partijen werken vaak samen, maar een kritische blik of het ontwikkelde algoritme geheel aansluit bij de taakuitvoering en de afdoende geconcretiseerde wensen en eisen van de opdrachtgever ontbreekt.

Het blijkt onduidelijker te worden wanneer externe organisaties onderdeel uitmaken van het proces en/of wanneer het gaat om verantwoordelijkheden op beleidsniveau of bestuurlijk niveau. Formeel is de betreffende minister verantwoordelijk voor de taakuitoefening door (een onderdeel van) de organisaties, en dus ook voor de inzet daarbij van algoritmes. Deze bevoegdheid is in de praktijk gemandateerd aan een functionaris in de lijn, zoals een algemeen directeur. Het is vaak niet eenduidig en vastgesteld welke directeur of bestuurder na implementatie de verantwoordelijkheid heeft voor het algoritme: de ontwikkelorganisatie (opdrachtnemer) of de proceseigenaar (opdrachtgever). Door niet expliciet vast te stellen wie waarvoor verantwoordelijk is, bestaat het risico dat uiteindelijk niemand verantwoordelijkheid neemt voor een algoritme en expliciete keuzes en beslissingen uitblijven, wat een negatief gevolg kan hebben voor de kwaliteit van de beheersing.

De verantwoordings- en rapportagelijnen dienen dan ook nog te worden vastgesteld. De controle op de ontwikkeling en toepassing van algoritmes is nog niet bepaald. Bij de ontwikkeling is de betrokkenheid van de tweede lijn, bestaande uit risico, controle en compliance toezichtfuncties die het verantwoordelijk management kunnen ondersteunen en adviseren, verschillend. Bij de meeste onderzochte algoritmes zijn (privacy) juristen geraadpleegd, interne adviseurs op het gebied van informatiebeveiliging spelen een kleinere tot geen rol.

Aanbeveling

Geef in een handreiking aan door wie en hoe verantwoording over de inzet en het functioneren van het algoritme afgelegd dient te worden. Tevens is het goed om aan te geven door wie controle wordt uitgevoerd of het algoritme voldoet aan relevante wet- en regelgeving en (interne) richtlijnen (bijvoorbeeld conform het 'Three Lines Model'), in aansluiting op de aanbeveling in paragraaf 3.3.

3.6 Beperkt inzicht in de wijze van beheersing van General IT Controls

Een algoritmes is geen op zichzelf staande applicatie. Het is een code die vaak op een computer meedraait of in andere programma's verwerkt is. Een goede

beheersing van de General IT Controls (GITC) is mede van belang voor het juist functioneren van een applicatie, alsmede voor de onderliggende systeemlagen (IT-omgeving). Deze set aan maatregelen moeten borgen dat de applicaties en systemen rond het algoritme en daarmee de uitkomsten betrouwbaar (tot stand gekomen) zijn. We hebben niet bij alle organisaties onderzoek uitgevoerd naar de GITC, omdat we hierin weinig risico's voor de betreffende (context van de) algoritmes zagen. Daar waar onderzoek naar GITC is gedaan, zijnde twee organisaties, is vanwege de beperkte verkregen informatie de wijze van beheersing niet inzichtelijk.

Aanbeveling

Benadruk het belang van het inzicht in de mate van beheersing van de GITC, met name indien de impact van het algoritme daarom vraagt, en het treffen van maatregelen indien nodig.

4 Privacy en ethiek

Dit hoofdstuk gaat in op de wijze waarop bij het gebruik van het algoritme invulling is gegeven aan ethische en privacy principes (deelvraag 2 en 3) om een verantwoorde inzet van algoritmes te borgen.

4.1 Meer aandacht voor zorgvuldige uitvoering van DPIA proces is nodig

Privacybeleid is doorgaans bij de onderdelen aanwezig. Omgang met algoritmes komt veelal niet specifiek terug in het privacybeleid. Waar het gaat om verwerking van persoonsgegevens, is de verwerkingsverantwoordelijke bekend. Ook weet men over het algemeen de (privacy) juristen te vinden.

De onderzochte algoritmes zijn niet alle voorzien van een DPIA. Als reden is aangegeven dat geen persoonsgegevens worden verwerkt of het algoritme onderdeel is van een groter geheel waar een DPIA voor zou moeten zijn/komen. De toelichting om geen DPIA uit te voeren is veelal niet vastgesteld. Daar waar wel een DPIA is uitgevoerd, is de samenhang tussen de uitwerking van risico's en maatregelen nog beperkt. Tot slot zien wij dat organisaties wachten met de beoordeling van risico's en de toetsing van randvoorwaarden totdat een algoritme formeel in gebruik wordt genomen. Het risico is dat in de pilotfase een organisatie privacy gerelateerde risico's niet juist adresseert en de gemaakte keuzes bij nader inzien niet goed genoeg zijn onderbouwd en opgelost dienen te worden om te voldoen aan de Algemene verordening gegevensbescherming (AVG).

Wij merken op dat organisaties soms nog zoekende zijn of een specifieke DPIA nodig is voor een algoritme, of dat deze betrekking moet hebben op het proces waar het algoritme onderdeel van uitmaakt. Bij één organisatie hebben wij gezien dat een DPIA is opgesteld voor het overkoepelende proces waarbij aandacht is besteed aan risico's en maatregelen gericht op het algoritme. Actualisatie van een bestaande DPIA van het proces is in bepaalde situaties een praktische werkwijze.

4.2 Noodzaak van verwerking van persoonsgegevens is doorgaans bepaald

Bij de ontwikkeling van algoritmes is aandacht geweest voor data-minimalisatie. Organisaties trachten niet meer gegevens dan noodzakelijk te verwerken middels het algoritme en hebben oog voor bezwaren vanuit privacy en ethisch oogpunt. Toepassing van pseudonimisering en aandacht voor en gebruik van ethisch gevoelige variabelen zijn wij als voorbeelden tegengekomen. Een punt van verbetering is de vastlegging van het besluit om bepaalde data te pseudonimiseren of uit te sluiten vanuit privacy en/of ethische overwegingen. Dit heeft een nauwe samenhang met onze bevinding over gestructureerde vastlegging van risico's, afwegingen en getroffen maatregelen in onder andere paragraaf 3.3.

4.3 Transparantie, bewaartermijnen en analyse van signalen van bias zijn nog aandachtspunten

Een aantal aspecten omtrent privacy en ethiek hebben bij de onderzochte algoritmes minder aandacht gekregen, wat gezien de status van het algoritme en de context/impact tot op zekere hoogte begrijpelijk is. Omdat deze aspecten in een andere of latere situatie belangrijk kunnen zijn, gaan we er hieronder op in.

Allereerst merken we op dat de transparantie over het gebruik van het algoritme richting publiek nog niet altijd aanwezig is. Dit hangt grotendeels af van het feit dat er geen sprake is van verwerking van persoonsgegevens, het algoritme een geringe

impact heeft en/of in ontwikkeling is. Desalniettemin is transparantie van belang op het moment dat persoonsgegevens worden verwerkt, wat ook tijdens de ontwikkeling (in pilotfase) het geval kan zijn. Een organisatie die reeds gebruik maakt van een algoritme waarbij persoonsgegevens worden verwerkt, maakt dit duidelijk kenbaar de eigen website en kan als good practice worden gezien.

Naast dat organisaties betrokkenen heldere informatie moeten geven over het gebruik van algoritmes, moeten ze ook gehoor geven aan mensen die een beroep doen op hun privacyrechten. De onderzochte organisaties verwijzen veelal naar het bestaande reguliere proces dat hiervoor bestaat. Waar het gaat om klachten van betrokkenen, merken we op dat nog niet altijd goed is nagedacht over de analyse van deze klachten, specifiek gericht op het functioneren van het algoritme, en de opvolging hiervan. Dit geldt ook voor de omgang met signalen vanuit de eindgebruiker over mogelijke bias of discriminatie door algoritmes. Het is van belang dat verzoeken, klachten en andere signalen terechtkomen bij de verantwoordelijke afdeling, zodat uitvoering en eventuele bijsturing kan plaatsvinden.

Tot slot zien we dat bewaartermijnen van de gebruikte data vaak bepaald zijn, maar dit voor de bewaartermijnen van de uitkomsten (output) van het algoritme nog niet altijd het geval is. Het gaat er dan om hoe lang bijvoorbeeld een lijst met personen of objecten, waarover de gebruiker uiteindelijk een besluit moet nemen, bewaard én gebruikt mag worden. Hiervoor dienen procedures, zoals voor de vernietiging van de uitkomsten, opgesteld en in gebruik te zijn om een juiste en rechtmatige verwerking te borgen.

4.4 Beoordeling door de organisaties van de impact en kans op discriminatie kan scherper

De onderzochte algoritmes leiden niet tot geautomatiseerde besluitvorming zonder menselijke tussenkomst. Een diepgaande analyse van de impact van het algoritme in brede zin is soms nog beperkt. Elk algoritme kent beperkingen die kunnen uitmonden in onjuiste of onvolledige output. Er lijkt soms te weinig bij de positieve en negatieve impact van een algoritme op de taakuitvoering en het uiteindelijke effect op de burgers, bedrijven en maatschappij te worden stilgestaan; de impact is dan ook niet altijd vastgelegd.

Bij algoritmes waarin persoonsgegevens worden verwerkt heeft het voorkomen van discriminatie aandacht. Wel is meer aandacht gewenst voor monitoring van mogelijke discriminatie. Ook hier geldt dat een bewuste afweging om al dan niet op bepaalde variabelen te monitoren vaak ontbreekt.

Aanbeveling

Geef in een handreiking aan wanneer, hoe en door wie risico’s ten aanzien van privacy en ethiek gestructureerd geanalyseerd en vastgelegd kunnen worden. Duid daarbij hoe het gebruik van algoritmes bij een werkproces kan worden opgenomen in een DPIA en hoe andere, reeds beschikbare instrumenten kunnen worden ingezet. Het is daarnaast aan te bevelen om de rol van functionarissen (tweede lijn), zoals privacy en security officers, te benoemen bij de toets of algoritmes juist en rechtmatig worden ingezet.

5 Model- en datakwaliteit

In dit hoofdstuk beschrijven we de wijze waarop de model- en datakwaliteit van de algoritmes is geborgd (deelvraag 5 en 6) door de onderzochte organisaties. De onderstaande bevindingen hebben hoofdzakelijk betrekking op de algoritmes die door de organisaties binnen het Rijk zelf zijn ontwikkeld. Bij het algoritme dat is ingekocht bij een externe partij is gering inzicht verkregen in het model en de data.

5.1 Modelkeuzes zijn onderbouwd en peer review hierop vindt in diverse vormen plaats

In veel gevallen is bij de start van het ontwikkeltraject een intakeformulier gehanteerd, waarin onder andere het beoogde doel van het algoritme beschreven staat. De ontwikkelaars konden voor elk onderzocht algoritme in de interviews onderbouwen welke overwegingen zijn gemaakt in de keuze voor het gebruikte modeltype en de onderliggende (hyper)parameters. In enkele gevallen is een dergelijke onderbouwing deels ook opgenomen in documentatie over het algoritme. Zoals aangegeven in paragraaf 3.3 valt echter over de gehele lijn nog verbetering te behalen omtrent deze vastlegging.

In de interviews en documentatie zijn diverse vormen van peer review aan de orde gekomen. Vaak werken meerdere otnwikkelaars samen in projecten, waardoor zij op de hoogte zijn van elkaars werk. In sommige gevallen wordt een peer review afgedwongen in een versiebeheerprogramma waarbij een checklist wordt gehanteerd met punten waar de code op moet worden gecontroleerd.

Onderbouwing van keuzes in documentatie en (afgedwongen) peer review volgens een vast sjabloon kunnen het risico beperken dat de keuze voor modeltype en (hyper)parameters niet optimaal is of dat kennis verloren gaat door uitdiensttreding van betrokkenen.

Aanbeveling

Overweeg of het wenselijk is om een baseline op te stellen waar peer reviews minimaal aan moeten voldoen.

5.2 Nauwkeurigheid en robuustheid van de modellen zijn volgens gangbare technieken beoordeeld door de organisaties in de ontwikkelfase

Uit de interviews en documentatie is gebleken dat over het algemeen controles op nauwkeurigheid en robuustheid van de modellen zijn toegepast in de ontwikkelfase. Bij de onderzochte *supervised learning* modellen is de data verdeeld in train-, validatie- en testsets, om nauwkeurigheidsmaten zoals *precision, recall* en *F1-scores* te berekenen. Ook is in sommige gevallen *cross validation* toegepast als techniek om een robuust model te verkrijgen. Met de benoemde waarden en techniek is een beeld verkregen van de prestatie van de modellen qua nauwkeurigheid en robuustheid. Daar waar elementen van het bovengenoemde achterwege gelaten zijn voor bepaalde algoritmes, is de keuze hiervoor toegelicht. Dit komt in de regel voort uit een afweging tussen kosten (vooral in tijd) en de te behalen winst in (zekerheid over) de prestatie van het model.

Nauwkeurigheid en robuustheid van de modellen worden wel gemeten, maar zoals ook benoemd in hoofdstuk 3 zijn slechts in beperkte mate acceptatiecriteria aangetroffen waarbij minimale waarden zijn gesteld aan bijvoorbeeld de

eerdergenoemde nauwkeurigheidsmaten. Daar waar wel criteria zijn aangetroffen is alleen gesteld dat deze waarden minimaal gelijk moeten zijn aan de behaalde waarden van de zogeheten 'Proof of concept'. De criteria zijn niet in verband gebracht met de functionele eisen en volgend uit een risicoafweging van de gebruikers van het algoritme of andere belanghebbenden.

Voor alle algoritmes is er sprake van handmatige verwerking door een medewerker, waardoor fouten of anderszins opvallende uitkomsten gesignaleerd kunnen worden. Actieve monitoring op de prestatie van het model door de tijd heen is echter op dit moment nergens ingericht. Dit sluit aan bij paragraaf 3.2, en hangt samen met het ontbreken van acceptatiecriteria waarop gemonitord zou moeten worden.

Aanbeveling

Overweeg om richtlijnen op te (laten) stellen waarmee een gestructureerde afweging vastgelegd kan worden welke technieken wel of niet worden toegepast voor het verkrijgen en beoordelen van nauwkeurigheid en robuustheid, gekoppeld aan de functionele eisen van het algoritme en een risicoanalyse. Daarnaast bevelen wij aan om te sturen op een werkwijze waarbij elk algoritme acceptatiecriteria kent, die in overleg tussen opdrachtgever, eindgebruiker en ontwikkelaar zijn vastgesteld.

5.3 Er is oog voor representativiteit van de gebruikte data

Uit de deelonderzoeken komt naar voren dat de representativiteit van de gebruikte inputdatasets voor het beoogde doel van het algoritme aandacht krijgt tijdens het ontwikkelproces en het gebruik van algoritmes. De organisaties hebben onderbouwd waarom zij de gebruikte train-, validatie- en testsets representatief achten. Een uitgewerkte analyse van meewegende aspecten hierin is vaak niet vastgelegd. Bij *unsupervised learning* is geen sprake van een train-, validatie- en testset. Er worden tijdens het ontwikkelproces echter wel hyperparameters bepaald en de prestatie van het algoritme wordt beoordeeld voor een specifieke dataset. Deze dataset moet daarom als trainingsdata worden gezien en dient representatief te zijn.

Bescherming van subpopulaties speelt bij de meerderheid van de onderzochte algoritmes geen rol omdat er in deze gevallen geen sprake is van uitkomsten die betrekking hebben op een individu of een groep mensen. Bij een algoritme waarvoor dit wel het geval is, is een good practice aangetroffen waarbij specifieke monitoring plaatsvindt op uitkomstbias voor bepaalde variabelen.

Aanbeveling

Vraag aandacht voor het gestructureerd onderbouwen van de representativiteit van de gebruikte data. Dit geldt ook indien een *unsupervised learning* techniek wordt toegepast.

5.4 Controles op juistheid, volledigheid en consistente codering van de data vinden plaats

In de eerder benoemde intakeformulieren en in documentatie wordt meestal veel toelichting gegeven op de gebruikte datasets en de bronnen waar deze van afkomstig zijn. Ontwikkelaars hebben in alle gevallen samengewerkt met de business, te weten opdrachtgevers en eindgebruikers van het algoritme, om domeinkennis te betrekken in het maken van keuzes over de te gebruiken variabelen en geprogrammeerde verwerkingsstappen (pre-processing). In sommige, maar niet alle gevallen zijn in de documentatie ook de resultaten van een zogeheten Exploratory Data Analysis (EDA) opgenomen, waarmee inzicht wordt gegeven in bepaalde eigenschappen van de dataset.

In de meeste gevallen zijn in de programmatuur controles opgenomen om de juistheid, volledigheid en consistente codering van de inputdata te beoordelen. Vaak is aangegeven dat de dataverwerking automatisch stopt als data bijvoorbeeld in het verkeerde format wordt aangeboden. Mogelijkheden zoals verbandscontroles, checks op *parsing errors* en lijncontroles zijn niet altijd uitputtend toegepast. Net als voor de keuzes in toegepaste nauwkeurigheid- en robuustheidsmaten is dit ook gebaseerd op afwegingen tussen kosten en de te behalen winst (in zekerheid over juistheid, volledigheid en consistentie van de data).

Aanbeveling

Overweeg aansturing gericht op gestructureerde vastlegging van keuzes over het wel of niet toepassen van controles op de juistheid, volledigheid en consistentie van de data.

5.5 De benodigde elementen om reproduceerbaarheid te borgen zijn veelal aanwezig

Aandacht voor reproduceerbaarheid van data en het model is op verschillende manieren aanwezig. Bewaartermijnen van (meta)data, het opslaan van de broncode in een versiebeheersysteem en het documenteren van gemaakte keuzes zijn elementen die hieraan bijdragen. Het expliciet opslaan van de output van een algoritme samen met de bijbehorende inputdata is nergens aangetroffen.

Bij een deelonderzoek is een good practice aangetroffen waarbij een applicatie is ingericht om op gestructureerde wijze (meta)data te verzamelen en op te slaan specifiek met het oog op de reproduceerbaarheid.

Aanbeveling

Stuur aan op een gestructureerde uitwerking van de mate waarin reproduceerbaarheid van model en data mogelijk moet zijn. Bepaal welke van de benoemde elementen minimaal geborgd moeten zijn, afhankelijk van de impact van het algoritme en gekoppeld aan een risicoanalyse.

6 Verantwoording onderzoek

6.1 Werkzaamheden en afbakening

De objecten van onderzoek betroffen de algoritmes van de geselecteerde organisaties. Daarbij is onderzocht op welke wijze de organisaties bij het gebruik van hun algoritme invulling hebben gegeven aan de volgende onderwerpen: governance, privacy, ethiek, GITC, input-, model- en outputkwaliteit.

In drie deelonderzoeken zijn de GITC buiten beschouwing gelaten, omdat dit voor de betreffende algoritmes niet relevant of haalbaar bleek. Dit is tussentijds afgestemd met de opdrachtgever alsmede het feit dat we niet alle onderdelen in opzet én bestaan hebben kunnen onderzoeken. De 'libraries' waarin voorgeprogrammeerde functies staan, vallen buiten de scope van dit onderzoek. Tevens is geen code-review gedaan en hebben we niet alle normen vanuit het ADR normenkader kunnen onderzoeken. De belangrijkste reden was dat drie algoritmes nog niet in productie waren, waardoor de normen die hier betrekking op hebben 'niet van toepassing' waren.

Voor de deelonderzoeken hebben we gebruik gemaakt van het ADR normenkader. De ADR heeft een normenkader ontworpen om algoritmes te onderzoeken. Het kader is ontwikkeld met behulp van nationale en internationale richtlijnen en rapporten, o.a. de (concept) richtlijnen van het ministerie van JenV, de *AI impact assessment* van het ECP, de *ethics guidelines for trustworthy AI* van de EC en het *Privacy Control Framework* van NOREA. Daarnaast is het kader afgestemd met (concept) toetsingskaders van andere partijen. De deelvragen in het onderzoek sluiten aan bij de onderwerpen in het normenkader.

Om te komen tot beantwoording van de onderzoeksvragen hebben we de benodigde gegevens verzameld in de periode augustus 2021 t/m februari 2022 middels het doornemen van data en van scripts, het bestuderen van relevante (kaderstellende) documenten en interviews met betrokkenen. Per deelonderzoek hebben we interviews gehouden met personen die vanuit de opdrachtgevende partij een rol spelen bij de governance, besluitvorming over ethische principes en/of de invulling van de privacyaspecten rondom het algoritme. Daarnaast is gesproken met de ontwikkelaars die aan het algoritme werken. De eindgebruikers van het algoritme, de medewerkers die de uitkomsten gebruiken voor hun werkzaamheden, zijn waar mogelijk betrokken bij het onderzoek.

De normen in het kader zijn voornamelijk onderzocht op opzet en niet op het bestaan in de praktijk. Tijdens het onderzoek hebben we geconstateerd dat we de normen in opzet goed konden onderzoeken maar dat het bestaan voor veel normen een onevenredig hoge controlelast op zou leveren voor de organisatie ten opzichte van de meerwaarde die het zou bieden in de bevindingen. Derhalve hebben we dit niet voortgezet waarmee is afgeweken van de opdrachtbevestiging. Dit is tussentijds afgestemd met de opdrachtgever.

De bevindingen in dit rapport geven inzicht in de wijze waarop de organisatie invulling geven aan de relevante aspecten/onderwerpen van een algoritme en gaan niet over de (juistheid van) gemaakte (ethische) keuzes t.a.v. het algoritme.

De bevindingen in dit rapport zijn afgestemd middels hoor en wederhoor.

Hiermee zijn de overeengekomen werkzaamheden uitgevoerd, met inachtneming van de bovengemelde tussentijdse inperking, conform de opdrachtbevestiging van dd 14 juni 2021 kenmerk 2021-0000087510.

6.2 Gehanteerde Standaard

Deze opdracht is uitgevoerd in overeenstemming met de Internationale Standaarden voor de Beroepsuitoefening van Internal Auditing.

Met dit onderzoek wordt geen zekerheid in de vorm van een oordeel of conclusie verschaft omdat het een onderzoeksopdracht betreft. Indien andere (aanvullende) werkzaamheden of een assurance-opdracht zouden zijn uitgevoerd, zouden wellicht andere onderwerpen zijn geconstateerd en gerapporteerd.

De opdracht is uitgevoerd conform de algemene uitgangspunten voor de uitoefening van de interne auditfunctie bij de rijksdienst. Daarbij hoort ook een stelsel van kwaliteitsborging. Een onderdeel daarvan is dat er een onafhankelijke kwaliteitstoetsing heeft plaatsgevonden op deze onderzoeksopdracht.

6.3 Verspreiding rapport

De opdrachtgever, CIO Rijk, is eigenaar van dit rapport.

De ADR is de interne auditdienst van het Rijk. Dit rapport is primair bestemd voor de opdrachtgever met wie wij deze opdracht zijn overeengekomen. Voor openbaarmaking door het opdrachtgevende ministerie van door de ADR aan dit ministerie uitgebrachte rapporten gelden de voorschriften uit de Wet open overheid. De minister van Financiën stuurt elk halfjaar een overzicht van door de ADR uitgebrachte rapporten naar de Tweede Kamer.

7 Ondertekening

Den Haag, 31 mei 2022

w/g

Projectleider
Auditdienst Rijk

Bijlage: Managementreactie

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

> Retouradres Postbus 20011 2500 EA Den Haag

Auditdienst Rijk
T.a.v. de heer
Postbus 20301
2500 EA Den Haag

Directoraat-Generaal
Directie CIO Rijk

Turfmarkt 147
Postbus 20011
2500 EA Den Haag

Contactpersoon

Kenmerk
2022-0000265958

Datum 30 mei 2022
Betreft Managementreactie ADR-onderzoek Toezicht op Algoritmen

Geachte heer

Ik dank de Auditdienst Rijk voor het onderzoeksrapport 'Beheersing algoritmes binnen het Rijk'. De daarin benoemde aanbevelingen zijn een belangrijke stap in het inrichten van het CIO-toezicht op algoritmen binnen het Rijk. Een belangrijke aanleiding voor CIO Rijk voor het verstrekken van deze opdracht aan de ADR was de constatering van de Algemene Rekenkamer[1] dat het nog aan structurele aandacht ontbreekt voor het toezicht op algoritmen. Doel van dit onderzoek was om inzicht te krijgen waar te beginnen met het inrichten van het CIO-toezicht op de ontwikkeling en inzet van algoritmen.

Voor het onderzoek heeft u ook een normenkader ontwikkeld dat gedurende het onderzoek zijn waarde heeft bewezen. In de rode draden rapportage geeft u op stelselniveau vier aanbevelingen voor de aanpak en prioriteiten om structurele aandacht voor algoritmes te organiseren.
Tijdens het opstellen van het rapport werkte u ook nauw samen met diverse onderdelen van de rijksoverheid. Dit resulteerde in vijf deelrapporten met aanbevelingen voor de betrokken organisaties; deze rapporten heeft u met de betrokken organisaties afgestemd. In deze managementreactie richt ik mij daarom op het normenkader en de rode dradenrapportage.

De beheersing van algoritmen is van groot belang en het gebruik van algoritmen in de samenleving en door de overheid groeit snel. Deze ontwikkeling en de grote impact ervan zijn ook beschreven door de Wetenschappelijk Raad voor het Regeringsbeleid[2]. Tegelijk zijn er risico's verbonden aan het gebruik van algoritmen door de overheid. Deze risico's, op bijvoorbeeld discriminatie, kwamen al naar voren in de rapportage van de Parlementaire ondervragingscommissie[3] Kinderopvangtoeslag. In de Tweede Kamer is hierna nog uitgebreid gesproken over hoe het risico op discriminatie met algoritmen kan worden beperkt en is de

[1] Aandacht voor algoritmes, Algemene Rekenkamer, januari 2021
[2] Opgave AI. De volgende systeemtechnologie, Wetenschappelijke Raad voor het Regeringsbeleid, november 2021
[3] Tweede Kamer, 2020/2021, 35 510 nr. 3

Directoraat-Generaal
Overheidsorganisatie
Directie CIO Rijk

Datum
30 mei 2022

Kenmerk
2022-0000265958

Motie Marijnissen[4] aangenomen om dit te beperken. Met de moties Klaver[5] en Dassen[6] wordt om een register voor algoritmen gevraagd. Met de motie Bouchallikht en Dekker-Abdulaziz[7] wordt gevraagd om verplicht gebruik van impact assessments voor mensenrechten en algoritmen. Met de motie Van Baarle[8] wordt aandacht gevraagd voor de inzet van handreikingen om discriminatie tegen te gaan.
Hieronder treft u mijn reactie per aanbeveling uit uw rapport. De managementreactie sluit ik af met een nawoord.

Reactie op uw aanbevelingen

1. Geef in een handreiking aan wanneer en hoe risicoanalyses rond het algoritme kunnen worden uitgevoerd, o.a. volgend op de wensen en eisen uit de business.
   Voor de Rijksoverheid zijn goede risicoanalyses rond algoritmen van groot belang. Het idee om dit in een handreiking te verwerken neem ik over. Het gebruik hiervan kan echter niet vrijblijvend zijn, daarom spreek ik liever van een beleidskader. Bij de uitwerking hiervan betrek ik het normenkader voor algoritmen dat u in dit onderzoek heeft gehanteerd en nu samen met de Algemene Rekenkamer verder ontwikkelt. Dit normenkader benoemt de verschillende risico's rondom ontwikkeling en gebruik van algoritmen. Het gebruik van dit normenkader zorgt voor de verbinding van de verantwoordelijkheden van de opdrachtgever- en het CIO-toezicht met het derdelijns toezicht.

   Deze handreiking zal worden opgenomen in de bestaande systematiek voor de beoordeling van (grote) ICT-projecten en opgenomen worden in het bijbehorende normenkader. In de te ontwikkelen handreiking wordt ook beschreven welke bestaande instrumenten daarbij kunnen worden gehanteerd, zoals de richtlijnen voor het toepassen van algoritmen[9], de Impact Assesment Mensenrechten en Algoritmen (IAMA), de handreiking non-discriminatie by design bij ontwikkeltrajecten, en aanvullende inkoopvoorwaarden voor leveranciers van systemen die algoritmen bevatten. De handreiking benoemt wanneer en op welke wijze dit type instrumenten kan worden ingezet. Ik ga deze handreiking in nauwe samenwerking met CIO' s binnen de rijksoverheid opstellen. Hierbij blijf ik graag een beroep doen op uw kennis en ervaring.

2. Geef in een handreiking aan wanneer en door wie risico's ten aanzien van privacy en ethiek bij algoritmes geanalyseerd en beheerst kunnen worden.
   Deze aanbeveling neem ik over. Ik verwerk die in de hiervoor al benoemde handreiking en in samenhang met de daar beschreven instrumenten.

---

4 Tweede Kamer, 2020/2021, 35 510 nr. 21
5 Tweede Kamer, 2020/2021, 35 510 nr. 16
6 Tweede Kamer, 2020/2021, 35 925 nr. 26
7 Tweede Kamer, 2021/2022, 26 643 nr. 835
8 Tweede Kamer, 2020/2021, 35 925 VII nr. 39
9 Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses. | Richtlijn | Rijksoverheid.nl

Directoraat-Generaal
Overheidsorganisatie
Directie CIO Rijk

Datum
30 mei 2022

Kenmerk
2022-0000265958

3. Overweeg een gerichte aanpak voor een eenduidige methode waarmee de kwaliteit van het model (nauwkeurigheid en robuustheid) en de data (representativiteit en juistheid/volledigheid/consistentie) onderbouwd kan worden.
   Het werken aan kwaliteit van data en een goed gebruik hiervan is inderdaad belangrijk. Met het inrichten en versterken van de rol van de Chief Data Officer zorg ik voor het eenduidig beleggen van de naleving op de regelgeving van de algoritmen. Hiernaast is het van belang om het risico van een mogelijke bias in data te beperken, wat ik wil doen met de inzet van de IAMA.

4. Geef in de handreiking aan:
   a. Wanneer en door wie de controle wordt uitgevoerd of het algoritme voldoet aan relevante wet- en regelgeving en (interne) richtlijnen.
   b. Wanneer en aan de hand van welke vooropgestelde criteria (o.a. acceptatie- en prestatiecriteria) het algoritme gemonitord en geëvalueerd kan worden.
   c. Door wie en hoe verantwoording over de inzet en het gebruik van het algoritme afgelegd dient te worden.

   Deze aanbeveling neem ik over. Ik verwerk de deeladviezen in de hiervoor al benoemde handreiking en in samenhang met de daar beschreven instrumenten.

Nawoord
Het rapport biedt een goede basis om verder aan de slag te gaan met het CIO-toezicht op de ontwikkeling en inzet van algoritmen bij het Rijk.
Daarvoor stel ik een plan van aanpak op dat na de zomer van 2022 afgestemd wordt met het CIO-beraad. Dit plan sluit aan bij de I-strategie Rijk 2021 - 2025 waar het belang van toezicht op algoritmen al is benoemd. Het plan van aanpak zal zich richten op het opstellen en implementeren van een beleidskader voor algoritmen. We starten nu al met een verkenning hiervan. Hierbij maken we graag weer gebruik van de kennis en ervaring van de ADR. Ook betrekken we hierbij de kennis en ervaring van de diverse onderdelen van de rijksoverheid.

Met dit rapport, het normenkader van de ADR en het toetsingskader van de Rekenkamer ligt er een solide basis om de komende periode het CIO-toezicht op algoritmen voor de rijksoverheid in te richten.

Met vriendelijke groet,

Auditdienst Rijk
Postbus 20201
2500 EE Den Haag
(070) 342 77 00