BIJLAGE waarborgenkader Douane

Governance					
	Beheersdoelstelling	De doelstelling en het gebruik van het algoritme is conform wet- en regelgeving en intern beleid/richtlijn(en) omtrent algoritmes. De taken, bevoegdheden en verantwoordelijkheden rond het algoritme zijn helder en sturing, beheersing en verantwoording vindt plaats. De risico's zijn geanalyseerd en zijn geaccepteerd of gemitigeerd met maatregelen. De organisatie monitort en evalueert periodiek de naleving van de maatregelen en effectiviteit en functionaliteit van het algoritme, met aandacht voor mogelijke ongewenste effecten.			
	Risico	Het gebruik van het algoritme is niet conform wet- en regelgeving en/of intern beleid. De sturing, beheersing en verantwoording t.a.v. het algoritme is beperkt: de doelstelling en beschrijving van het algoritme inclusief onderbouwing zijn niet aanwezig is. Taken, bevoegdheden en verantwoordelijkheden zijn niet helder of adequaat belegd, risico's worden niet geanalyseerd en gemanaged en de effectiviteit en functionaliteit van het algoritme worden niet gemonitord en geëvalueerd.			
Nr.	**Deelgebied**	**Norm**	**Risico**	**Suggested test steps/evidence**	**Toelichting norm (eventueel)**
GOV.1	Beleid	Een beleid of richtlijn waarin de voorwaarden aan de ontwikkeling en het gebruik van algoritmes zijn beschreven is aanwezig.	Het risico bestaat dat het algoritmes binnen de organisatie niet voldoen aan het interne beleid en de onderliggende wet- en regelgeving.	Beleidsdocument of richtlijn(en). Documentatie over vaststelling. Nagaan in interviews.	Het (intern) beleid of richtlijnen is vastgesteld door het hoogste management, wordt periodiek herzien en is bekend bij de betrokkenen.
GOV.2		Het is vastgesteld dat het algoritme voldoet aan het beleid of de richtlijn.	De inzet van het algoritme is niet juist, wat diverse risico's creëert.	Documentatie waaruit de vaststelling blijkt, bijv. een controle door 1e/2e lijn. Nagaan in interviews.	
GOV.3	Doelstelling	De doelstelling van het algoritme en op welke wijze dit bijdraagt aan de taakuitvoering van de organisatie is vastgesteld.	Zonder eenduidigheid over het doel is geen sturing op en verantwoording over het algoritme mogelijk.	Beschrijving algoritme in bijv. functionele en technische documentatie/ontwerp. Eventueel andere interne stukken.	
GOV.4		Een bewuste afweging is gemaakt of het algoritme het juiste middel is om de taakuitvoering op doelmatige wijze te realiseren.	Risico is dat de inzet van een algoritme als instrument niet nodig is bij de taakuitvoering, maar de inzet van een eenvoudigere techniek ook voldoende is. Dit voorkomt risico's die gepaard gaan met een algoritme.	Beschrijving algoritme in bijv. functionele en technische documentatie/ontwerp. Eventueel andere interne stukken.	Het is niet zondermeer nodig AI methoden in te zetten op data. Vaak zijn min of meer traditionele analysemethoden geschikt om de kwaliteit van bronnen te onderzoeken of om patronen te vinden. Dit moet en bewuste afweging zijn, ook met oog op de taakstelling.
GOV.5	Uitlegbaarheid (collegiaal)	Een actuele beschrijving van het algoritme is aanwezig. De beschrijving bevat de doelstelling(en) van het algoritme, het type algoritme, de gebruikte gegevens (incl. bron), (hyper)parameters en prestatiecriteria.	Gebrek aan een beschrijving zorgt intern voor onduidelijkheid en kan leiden tot interpretatieverschillen en fouten bij de ontwikkeling en het gebruik van het algoritme.	Beschrijving algoritme in bijv. functionele en technische documentatie/ontwerp. Bij voorkeur staan alle aspecten in één of beperkt aantal document(en). Bij betrokkenen nagaan of deze ook bekend is en het de laatste, actuele versie betreft.	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, ook rekening houdend met gaming the system.
GOV.6		De gemaakte overwegingen en keuzes bij het algoritme zijn vastgelegd.	Risico is dat overwegingen en keuzes niet helder zijn voor alle betrokkenen en daardoor onjuist worden toegepast.	Document waarin overwegingen en keuzes staan en zijn onderbouwd. Bijvoorbeeld in beschrijving of functionele/technische documentatie.	
GOV.7		Bij het (door)ontwikkelen van het algoritme zijn de relevante functionarissen en belanghebbenden betrokken.	Risico is dat kennis ontbreekt en het algoritme uiteindelijk niet rechtmatig en betrouwbaar is en bijdraagt aan het vastgestelde doel.	Nagaan of functionarissen en belanghebbenden in de organisatie die te maken hebben met het algoritme zijn geïnformeerd en betrokken, in interviews en uit interne documentatie.	Denk aan de eigenaar, ontwikkelaar en gebruiker.
GOV.8		De beschrijving is opgeslagen op een plaats waar de betrokken functionarissen volledig toegang/inzicht hebben.	Wanneer de beschrijving niet toegankelijk en inzichtelijk is, kunnen bij de overdracht ongemerkt en onbedoeld fouten / interpretatieverschillen ontstaan.	De beschrijving kan op diverse plekken opgeslagen zijn, zoals Git, samenwerkingsruimte (Sharepoint of MS teams), netwerkschijf of DMS. Het is van belang dat het op één plek staat, waar betrokkenen bij kunnen. Dit ook nagaan in interviews.	Teams en afdelingen moeten volledig toegang hebben tot de laatste versie van de beschrijving en andere relevante documentatie.
GOV.9	Gebruik uitkomsten algoritme	Het is vastgesteld aan welke interne en externe partijen de uitkomsten van het algoritme worden verstrekt en inzage wordt verleend.	Risico is dat de uitkomsten van het algoritme niet worden gebruikt waarvoor ze zijn bedoeld en onrechtmatig verwerkt worden.	Beschrijving algoritme in bijv. functionele en technische documentatie/ontwerp inclusief onderbouwing.	Uitkomsten zijn binnen overheidsdomein veelal risicoscores, maar kunnen ook besluiten zijn als de toezegging van een subsidie of toeslag. Het leidend principe is dat de

					uitkomsten zo min mogelijk worden verstrekt, alleen indien noodzakelijk. Dit geldt zeker wanneer persoonsgegevens worden verwerkt.
GOV.10		Het is vastgesteld dat de uitkomsten van het algoritme alleen worden gebruikt voor het betreffende doel en verstrekt aan de partijen waarvoor het bestemd is.	Risico is dat de uitkomsten van het algoritme niet worden gebruikt waarvoor ze zijn bedoeld en onrechtmatig verwerkt worden.	Documentatie waarin de ontvangers (intern en extern) zijn gedefinieerd en nagaan bij betrokkenen (bijv. gebruiker en eindverantwoordelijke) met wie de uitkomsten gedeeld worden.	
GOV.11		De bewaar- en gebruiktermijn van de uitkomsten van het algoritme voor de partijen zijn vastgesteld en worden nageleefd.	Risico is dat dat de uitkomsten langer dan noodzakelijk worden bewaard en gebruikt dan dat de wet- en regelgeving voorschrijft.	Beschrijving algoritme in bijv. functionele en technische documentatie/ontwerp.	De bewaar- en gebruiktermijn is gebaseerd op het intern beleid / richtlijnen, welke een vertaling vormen van wet- en regelgeving (bv. AVG en Archiefwet).
GOV.12	Wet- en regelgeving	Het is vastgesteld dat het algoritme voldoet aan de actuele en relevante wet- en regelgeving.	Het gebruik van het algoritme is onrechtmatig. Dit kan leiden tot schade voor individuen en de maatschappij, politiek-bestuurlijke en/of juridische maatregelen, verlies van vertrouwen en beschadiging van imago.	Interne toets (door bv. juridische afdeling/2LOD of 1LOD) of is voldaan aan juridische en beleidsmatige toetsingskaders.	Het betreft hier wet- en regelgeving die eisen stelt aan het gebruik van een algoritme en data of het domein/beleidsterrein. Bij de verwerking van persoonsgegevens speelt de AVG een belangrijke rol (zie 'Privacy'), maar bijv. ook Algemene beginselen van behoorlijk bestuur, Wet politiegegevens (Wpg) en Wet justitiële en strafvorderlijke gegevens (Wjsg), Archiefwet en Beveiligingsrichtlijnen (zie 'GITC') kunnen van toepassing zijn.
GOV.13	Taken, bevoegdheden, verantwoordelijkheden	De taken, bevoegdheden en verantwoordelijkheden rond het algoritme zijn vastgesteld en in praktijk toegepast.	Onduidelijkheid over taken, bevoegdheden en verantwoordelijkheden kan zorgen dat geen (effectieve) invulling wordt gegeven aan de wensen en eisen vanuit intern beleid en wet- en regelgeving en sturing en verantwoording ontbreekt.	Beschrijving van TBV's in een goverance-beschrijving. Documentatie waaruit toepassing in praktijk blijkt, o.a. in vastlegging formele besluiten of andere reguliere stukken/verslagen.	Het gaat om de TBV's van betrokkenen bij het ontwerp, ontwikkelen en gebruik van het algoritme, zoals: - eindverantwoordelijke (proceseigenaar) - gebruikersorganisatie - ontwikkelorganisatie - verwerkersorganisatie (verwerker) - beheerorganisatie - data-eigenaar - privacyofficer en/of jurist. Formeel is de betreffende minister eindverantwoordelijk voor de taakuitoefening binnen de Rijksoverheid. In de praktijk zal de verantwoordelijkheid zijn gemandateerd, bv. aan een DG of een directeur. Regie op en verantwoordelijkheid voor het ontwerp- en ontwikkelproces dient duidelijk te worden belegd (bij een team of persoon) en in het proces zelf (na implementatie) bij de (lijn)organisatie.
GOV.14		De eindverantwoordelijke heeft de ingebruikname van het algoritme geaccepteerd na een toets of het algoritme voldoet aan de vooraf opgestelde eisen en criteria.	Risico dat de ingebruikname niet door de bevoegde persoon verantwoord wordt en voldoet aan de gestelde eisen en criteria.	Documentatie van formele goedkeuring eindverantwoordelijke (opdrachtgever), en de acceptatietoets of -dossier.	De vooraf opgestelde eisen en criteria zijn bepaald in de opdrachtformulering en/of opgenomen in de beschrijving van het algoritme. De eindverantwoordelijke dient voor de implementatie te toetsen of hieraan wordt voldaan (en het algoritme valide en betrouwbaar is).
GOV.15		De eindverantwoordelijke legt verantwoording af over de inzet en ontwikkeling van het algoritme en de effectiviteit en functionaliteit van het algoritme.	Risico dat de inzet van het algoritme ongeoorloofd is doordat het niet door de bevoegde en verantwoordelijke funtionaris verantwoord wordt.	Documentatie van besluit en onderbouwing voor inzet algoritme (verantwoordingsdocument) en documentatie over gerealiseerde effectiviteit (rapportage).	De eindverantwoordelijke legt voorafgaand verantwoording af over de inzet van het algoritme en achteraf over de gerealiseerde effectiviteit en functionaliteit, wat uit een evaluatie moet blijken. De verantwoording gebeurt volgens het principe van comply or explain: organisaties moeten het intern beleid / richtlijn volgen, of nadrukkelijk uitleggen waarom zij ervan afwijken.
GOV.16		Bij uitbesteding van onderdelen of activiteiten met betrekking tot het algoritme aan externe partijen zijn afspraken gemaakt en vastgelegd om een te grote afhankelijkheid te voorkomen.	Risico dat een te grote afhankelijkheid (lock-in) ontstaat, waardoor overstappen niet (meer) mogelijk is en/of tot hoge kosten leidt.	Overeenkomst met externe partij. Door beide partijen ondertekend (conform TBV's). Maatregelen om een te grote afhankelijkheid te voorkomen zijn hierin beschreven. Eventueel is exitstrategie opgesteld.	Vendor lock-in maakt een organisatie afhankelijk van een externe partij (ontwikkelaar), omdat ze niet in staat is om van partij te veranderen zonder substantiële omschakelingskosten of ongemak. Naast afhankelijkheid van systemen gaat het ook om kennis die na de uitbesteding eventueel verloren gaat.
GOV.17	Deskundigheid	De vereiste deskundigheid, competenties en capaciteit is aanwezig in de organisatie.	De betrokken functionarissen bij het algoritme hebben onvoldoende deskundigheid en capaciteit, waardoor signalen over gebreken of biases in het algoritme gemist worden en risico's ontstaan.	Nagaan bij betrokkenen, evt. aangevuld met documentatie zoals personeels- en opleidingsplan met benodigde compententies en capaciteit of projectdocumentatie waarin aanwezige/benodigde	Het personeel (intern en extern) dat werkt met het algoritme moet kwalitatief en kwantitatief over voldoende deskundigheid en competenties beschikken. Dit moet in ieder geval blijken uit de risicoanalyse, waarbij de

				competenties en capaciteit van het algoritme is beschreven.	organisatie maatregelen neemt om eventuele gaten te overbruggen.
GOV.18	Risicomanagement	De risico's bij het gebruik van het algoritme worden periodiek geanalyseerd en gedocumenteerd.	Onbedoelde nadelige effecten van het algoritme ontstaan van het algoritme worden niet tijdig vastgesteld en adequaat geadresseerd (o.a. wanneer ontwerpers op afstand komen te staan).	Risicoanalyse-document van het algoritme en communicatie van geïnventariseerde risico's richting intern betrokkenen. Hierbij kan gebruik worden gemaakt van een DPIA (bij persoonsgegevens) of AIIA. Bij betrokkenen functionarissen nagaan welke rol zij hierbij hebben gespeeld. Analyse effectiviteit algoritme.	Risicoanalyses vinden, vooraf en tijdens het gebruik van het algoritme plaats (bij ontwikkel- en/of gebruikersorganisatie). Het gata o.a. om risico's m.b.t. fouten en ongewenste bias in algoritme of datasets, discriminerende- of stigmatiserende effecten en compliance risico's, maar ook risico's voor de organisatie (imago en bedrijfsvoering). De resultaten van de evaluatie van de effectiviteit en functionaliteit dient mee te worden genomen. De risicoanalyses zijn gecommuniceerd aan de eindverantwoordelijke.
GOV.19		De risico's zijn geaccepteerd of gemitigeerd met maatregelen. Deze maatregelen worden periodiek gemonitord en geëvalueerd.	Er wordt geen follow-up gegeven aan onderkende risico's.	Documentatie waaruit opvolging risicoanalyse blijkt, zoals risico-analyse(s) of rapportage(s) en reguliere stukken/verslagen. De mate waarin de risico's zijn geaccepteerd moet hierin naar voren komen, evenals de voorgenomen en (status van) getroffen maatregelen en de monitoring en evaluatie hiervan. Tevens nagaan bij betrokkenen.	Bij voorkeur in normale P&C-cylcus, waarbij de eindverantwoordelijke (proceseigenaar) actief betrokken is.
GOV.20	Evaluatie	Afspraken over de wijze waarop het algoritme wordt gemonitord en geëvalueerd zijn vastgelegd.	Gebrek aan afspraken kan leiden dat het algoritme niet of niet juist wordt gemonitord, waardoor mogelijke onbedoelde nadelinge effecten en andere gebreken niet tijdig worden vastgesteld.	Plan van aanpak voor evaluatie. Dit kan een los document zijn of onderdeel uitmaken van bijv. de beschrijving of risico-analyse.	Voor de ingebruikname is vastgelegd door wie, wanneer (frequentie) en hoe het algoritme gemonitord en geëvalueerd wordt. Dit moet geïntegreerd zijn in het risicomanagement.
GOV.21		De organisatie monitort en evalueert periodiek de effectiviteit en functionaliteit van het algoritme met aandacht voor mogelijke ongewenste effecten.	Signalen over gebreken of biases in het algoritme worden niet vastgesteld, waardoor het algoritme niet gecorrigeerd en beheerst wordt.	Evaluatie(s)/rapportage of andere interne documentatie hierover.	Naast de effectiviteit en functionaliteit moet gekeken worden of het algoritme geen geen bijkomende schade oplevert voor anderen (zie ook ethiek). De evaluatie dient met de eindverantwoordelijke en andere relevante betrokkenen gedeeld te worden.
GOV.22		De organisatie analyseert of (interne en externe) klachten en incidenten het gevolg kunnen zijn van het algoritme.	Signalen over gebreken of biases in het algoritme worden niet (als zodanig) vastgesteld, waardoor het algoritme niet gecorrigeerd wordt.	Opname analyse(resultaten) in evaluatie van het algoritme of in de beschrijving klachten- of incidentenproces. Eventueel mondelinge toelichting over wijze van de uitgevoerde analyse.	Burgers of andere belanghebbenden kunnen via reguliere processen een klacht of bezwaar indienen tegen de uitkomsten van het algoritme. Ook intern moet de mogelijk zijn om klachten of zorgen te kunnen uiten. Het moet ingeregeld zijn dat deze klachten worden meegenomen in de analyse. De analyseresultaten dienen te worden gedeeld met de eindverantwoordelijke en andere relevante betrokkenen.
Privacy					
	Beheers doelstelling	Het algoritme is in overeenkomst met het privacybeleid van de organisatie en de privacywetgeving (AVG) om te waarborgen dat de verwerking van gegevens op een rechtmatige wijze plaatsvindt. De kenmerken en rechtmatigheid van de gegevensverwerking zijn beoordeeld middels een DPIA. De organisatie heeft hierin tevens de privacyrisico's beoordeeld om te bepalen hoe deze, door het treffen van maatregelen, teruggebracht kunnen worden binnen de grenzen die de organisatie acceptabel vindt. De wijze waarop de persoonsgegevens worden verwerkt is voor het publiek en de betrokkene transparant en maakt het de betrokkene mogelijk zijn rechten uit te oefenen.			
	Risico	Het ontbreken van een privacybeleid leidt ertoe dat de organisatie rond het algoritme geen duidelijkheid heeft wat precies wordt verwacht. Privacyrisico's worden niet of niet tijdig gesignaleerd, waardoor de verwerking van de persoonsgegevens niet aan de privacywetgeving (AVG) voldoet en een grote(re) kans loopt op inbreuken op de beveiliging. Dit maakt de verwerking onrechtmatig en kan leiden tot schade voor de betrokkene.			
Nr.	**Deelgebied**	**Norm**	**Risico**	**Suggested test steps/evidence**	**Toelichting norm (eventueel)**
PRI.1	Privacybeleid	Het algoritme voldoet aan het privacybeleid.	Het algoritme niet voldoet aan het privacybeleid en de privacywetgeving, waardoor de verwerking met het algoritme onrechtmatig is.	Privacybeleid (organisatiebreed of specifiek voor onderdeel) en controle/ vaststelling dat het algoritme hieraan voldoet (bijv. door juridische afdeling).	De organisatie heeft een privacybeleid vastgesteld waarin de doelstellingen en verantwoordelijkheden m.b.t. privacy staan en is i.o.m. geaccepteerde privacyprincipes en de van toepassing zijnde wet- en regelgeving. Het privacybeleid is een concrete vertaalslag van de AVG-normen naar de gegevensverwerkingen van een organisatie. Normen uit de AVG herhalen is niet voldoende. Een privacybeleid is niet altijd verplicht volgens AP. Het privacybeleid is bij voorkeur één document, maar kan ook zijn vastgelegd in meerdere documenten.

PRI.2	Verwerkings-verantwoordelijkheid	De verwerkingsverantwoordelijke en verwerker van het algoritme zijn vastgesteld. Indien relevant is een verwerkersovereenkomst aanwezig.	Sturing op de inrichting van privacy en naleving is niet helder en zorgt dat het algoritme niet voldoet aan het privacybeleid en privacywetgeving.	DPIA en beschrijving algoritme. Eventueel kan gekeken worden naar benoeming in verwerkingsregister of verwerkersovereenkomst.	Een verwerker verwerkt persoonsgegevens in opdracht van een andere organisatie en is niet rechtstreeks onderworpen aan het gezag van de verwerkingsverantwoordelijke.
PRI.3	Data Protection Impact Assessment (DPIA)	Een DPIA is uitgevoerd op de gegevensverwerking met het algoritme. Het besluit om geen DPIA uit te voeren is onderbouwd.	Risico dat niet wordt voldoen aan wettelijke verplichting (AVG) met betrekking tot uitvoeren DPIA, waardoor de privacyrisico's en de te treffen maatregelen van de verwerking niet inzichtelijk zijn.	Uitgevoerde (en ondertekende) DPIA of onderbouwing (risicoafweging) waarom geen DPIA is uitgevoerd. Dit dient te zijn bekrachtigd door het management.	Data protection impact assessment (DPIA), in het Nederlands ook wel de gegevensbeschermings-effectbeoordeling (GEB), moet voorafgaand aan gegevensverwerkingen met een hoog privacyrisico worden uitgevoerd. Een DPIA is vrijwel altijd verplicht (zie AP). O.a. bij automatische verwerking (profilering) die leiden tot een besluit met gevolgen voor mensen, verwerkingen van bijzondere of strafrechtelijke gegevens of gegevens van mensen in een publiek toegankelijk gebied (bijv. cameratoezicht) op grote schaal. DPIA moet bij (grote) wijzigingen in de verwerking of bij langdurige verwerking herhaald worden. Bij de uitvoering dienen i.i.g. de FG/CISO betrokken te zijn.
PRI.4		Aan de voorgenomen maatregelen uit de DPIA is uitvoering gegeven.	Privacyrisico's bij de verwerking met het algoritme blijven bestaan.	DPIA-rapportage en document met opvolging uitkomsten (bijv. verwerkt in de beschrijving algoritme) dan wel plan van aanpak om opvolging te geven. Het kan ook meegenomen zijn in risico rapportages.	
PRI.5	Rechtmatige grondslag	De rechtmatige grondslag voor de verwerking van persoonsgegevens door het algoritme is vastgesteld.	De verwerking van de persoonsgegevens met het algoritme is onrechtmatig.	DPIA-rapportage, eventueel in beschrijving algoritme	
PRI.6	Doelbinding	Het is vastgesteld dat de verwerking van persoonsgevens met het algoritme verenigbaar is met het oorspronkelijke doel.	De verwerking voldoet niet aan doelbinding en is dus onrechtmatig.	DPIA-rapportage, eventueel in beschrijving algoritme	
PRI.7	Data-minimalisatie	Het is vastgesteld dat het algoritme niet meer persoonsgegevens verwerkt dan noodzakelijk.	De verwerkte gegevens zijn niet proportioneel en relevant in relatie tot het doel.	DPIA-rapportage, eventueel in beschrijving algoritme	
PRI.8	Bewaartermijn	De bewaartermijn van persoonsgegevens (inputvariabelen) en uitkomsten van het algortime is vastgesteld. Een procedure voor het tijdig vernietigen hiervan is vastgesteld.	De gegevens en uitkomsten worden langer bewaard dan toegestaan.	DPIA-rapportage en beschrijving algoritme Procedure voor vernietigen.	
PRI.9	Transparantie (uitlegbaarheid)	De betrokkenen worden geïnformeerd over de verwerking van persoonsgegevens door het algoritme en de verwachte gevolgen.	De betrokkenen zijn niet op de hoogte van de verwerking. Door gebrek aan transparantie kan geen verantwoording worden afgelegd en kan achteraf (bij fouten) tot kosten/schadeclaims leiden.	Privacystatement, evt. voorbeelden van communicatie en berichtgeving naar betrokkenen. Communicatie moet openbaar zijn, bijv. op de website.	
PRI.10	Rechten van betrokkenen	Het is vastgesteld dat de uitvoering van de rechten van betrokkenen is gewaarborgd bij het gebruik van het algoritme.	De betrokkenen kunnen hun rechten volgens de wet niet uitoefenen.	DPIA-rapportage en document met opvolging uitkomsten (bijv. verwerkt in de beschrijving algoritme) dan wel plan van aanpak om opvolging te geven. Nagaan bij betrokkenen (jurist).	
PRI.11	Besluitvorming	Het is vastgesteld dat de betrokkenen de mogelijkheid hebben om niet onderworpen te zijn aan geautomatiseerde besluitvorming door het algoritme zonder menselijke tussenkomst.	De verwerking voldoet niet aan de wettelijke verplichting AVG en het hanteren van de menselijke maat.	DPIA-rapportage, met onderbouwing (mogelijkheid) tot menselijke tussenkomst algoritme	
Ethiek					
	Beheersdoelstelling	De impact van het algoritme op burgers, bedrijven, de maatschappij en het milieu (en bijbehorende risico's) is beoordeeld en meegenomen in de gemaakte overwegingen en keuzes t.a.v. het algoritme. Het algoritme houdt rekening met diversiteit en discrimineert of stigmatiseert niet. Indien het algoritme een aanmerkelijke impact heeft op de besluitvorming, is betekenisvolle menselijke tussenkomst georganiseerd om risico's te beperken. De organisatie is transparant naar het publiek en belanghebbenden over het gebruik van een algoritme en kan een beslissing of besluit met het algoritme herroepen.			
	Risico	De impact van het algoritme is niet beoordeeld, waardoor overwegingen en keuzes niet of niet goed zijn onderbouwd. Het algoritme discrimineert of stigmatiseert (ongemerkt en onbedoeld) bepaalde personen, groepen of andere eenheden. Het publiek en belanghebbenden zijn niet op de hoogte van het gebruik van het algoritme. Dit gebrek aan transparantie ondermijnt het vertrouwen van de burger en mogelijkheid tot controle. Het algoritme maakt zo inbreuk op grondrechten en voldoet niet aan wet- en regelgeving. Dit kan achteraf leiden tot kosten/schadeclaims en bestuurlijke/politieke schade.			

Nr.	Deelgebied	Norm	Risico	Suggested test steps/evidence	Toelichting norm (eventueel)
ETH.1	Eerlijkheid (fairness)	De impact van het algoritme op burgers, bedrijven, de maatschappij en het milieu is geanalyseerd en beoordeeld.	De impact van het algoritme is niet helder en niet beoordeeld, waardoor afwegingen en keuzes mogelijk niet of niet juist gemaakt zijn.	Document met effectbeoordeling, bv. Artificial Intelligence Impact Assessment (AIIA). De AIIA is een breder risico-inschattingsinstrument dan de DPIA en richt zich op alle mogelijke ethische en juridische vraagstukken m.b.t. het algoritme. Resultaten uit DPIA zijn hierin bij voorkeur verwerkt. De beoordeling moet voorafgaand aan de ontwikkeling en ingebruikname zijn uitgevoerd. De afwegingen daarbij zijn gedocumenteerd.	De mate van impact verschilt per algoritme: een voorschrijvend algoritme (besluitvorming) heeft in de regel meer impact dan een beschrijvend algoritme. Algoritme die rechtsgevolgen of anderszins een aanmerkelijke impact hebben op burgers, bedrijven, of (groepen in) de maatchappij en milieu vragen hogere eisen. De afwegingen (o.a. tegen het algemeen of maatschappelijk belang) zijn gedocumenteerd.
ETH.2		De kans op discriminatie of stigmatisering door het algoritme is beoordeeld en geminimaliseerd.	Het algoritme schendt de grondrechten doordat het discrimineert of stigmatiseert.	Document met effectbeoordeling, bv. Artificial Intelligence Impact Assessment (AIIA). Ook methoden of mechanisme(s) om de kans te minimaliseren zijn gedocumenteerd in bijv. de beschrijving algoritme.	Het gaat hierbij om de beoordeling van de grondrechten. Komt ook terug bij andere onderdelen, waaronder Governance (risicomanagement) en Input-, Model- en Output-kwaliteit (daadwerkelijke minimalisatie).
ETH.3		Controlemechanismes die specifiek toetsen of het algoritme niet leidt tot discriminatie of stigmatisering zijn aanwezig.	Het algoritme schendt de grondrechten doordat het discrimineert of stigmatiseert.	Beschrijving van controlemechanisme(s).	Controlemechanismen moeten binnen de grenzen van de wet zijn. Soms kan het echter 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 laat controle met behulp van bijzondere persoonsgegevens, zoals gegevens over iemands ras, vooralsnog niet toe.
ETH.4	Respect voor de menselijke autonomie	Indien het algoritme leidt tot geautomatiseerde besluitvorming dan wel anderszins een aanmerkelijke impact heeft, is bepaald of betekenisvolle menselijke tussenkomst noodzakelijk of wenselijk is.	Gebruik van geautomatiseerde besluitvorming of gebrek aan mogelijke menselijke tussenkomst is niet toegestaan of onwenselijk.	Beschrijving algoritme (technische en functionele ontwerp/documentatie).	Het gaat hierbij om de mate van invloed en autonomie van het algoritme op het proces en de uitkomst zoals individuele besluitvorming. Ook bij voorspellende algoritmes die niet in besluitvorming uitmonden, zoals een risicotaxatie of beleid, maar wel een aanmerkelijke impact hebben, dient (betekenisvolle) menselijke tussenkomst organiseerd te worden. Het algoritme moet rekening houden met de "menselijke maat" (zie SyRi-arrest/ RvS). Menselijke tussenkomst is niet altijd nodig, zoals bij de administratiefrechtelijke afdoening van verkeersovertredingen, omdat de uitkomst volledig op regelgeving gebaseerd is.
ETH.5		De mogelijkheid tot betekenisvolle menselijke tussenkomst is beschreven en in het proces ingebed.	Onduidelijkheid over of gebrek aan menselijke tussenkomst kan grondrechten schenden.	Beschrijving algoritme (technische en functionele ontwerp/documentatie) en evt. beschrijving van het proces waarin het algoritme wordt gebruikt.	Het moet helder zijn wat het niveau is van menselijke controle en betrokkenheid, wie is de 'mens in controle' is en wat de momenten of instrumenten voor menselijk ingrijpen zijn.
ETH.6		Besluiten genomen door of aan de hand van het algoritme zijn herroepelijk. Een eenduidige herbeoordelingsprocedure is vastgesteld.	De beslissing of het besluit kan niet worden herzien, waardoor niet aan wettelijke verplichtingen wordt voldaan.	Herbeoordelingsprocedure van besluit algoritme.	Zie ook onderdeel Outputkwaliteit.
ETH.7	Transparantie en verantwoording	De organisatie deelt algemene informatie over het algoritme actief met het publiek, zoals het gebruik, de doelstelling(en) en eventuele consequenties van het algoritmEth.	Belanghebbenden weten niet dat gebruik wordt gemaakt van een algoritme. Door gebrek aan transparantie kan geen verantwoording worden afgelegd en kan achteraf (bij fouten) tot kosten/schadeclaims leiden.	Privacystatement; voorbeelden van communicatie en berichtgeving naar betrokkenen	Wanneer persoonsgegevens worden verwerkt, dient dit uitlegbaar te zijn voor collega's en vanuit de AVG voor de betrokkenen. Vanuit ethisch perspectief dient het algoritme ook transparant voor 'het publiek' te zijn (ongeacht het type gegevens), toegesneden op de specifieke kenmerken van algoritmische data-analyses (richtlijnen J&V). Dit kan ook ingegeven zijn vanuit de Wob / Awb. Het verschilt per algoritme en organisatie hoeveel transparantie gegeven wordt, mede gezien het risico op *gaming the system*.Over het algemeen geldt dat naarmate de impact groter is, transparantie belangrijker wordt.
ETH.8		De informatie over het algoritme is beknopt, begrijpelijk en gemakkelijk toegankelijk.	Het is voor belanghebbenden niet duidelijk dat zij te maken hebben met een algoritme en welke gevolgen dit voor hun heeft.	Privacystatement; voorbeelden van communicatie en berichtgeving naar betrokkenen	
Informatiebeveiliging: GITC					
	Beheersdoelstelling	De beheersingsmaatregelen waarborgen dat ongeautoriseerde wijzigingen in het algoritme worden voorkomen. Het wachtwoordbeheer is interactief en moet bewerkstelligen dat wachtwoorden van geschikte kwaliteit worden gekozen om het risico op misbruik te voorkomen. De beheersingsmaatregelen waarborgen dat de logische toegang tot het algoritme is beperkt tot daartoe bevoegde personen. De beheersingsmaatregelen waarborgen dat het algoritme is beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen. De beheersingsmaatregelen waarborgen dat back-ups in overeenstemming met het back-upbeleid worden gemaakt en dat het algoritme bij gegevensverlies hersteld kan worden.			

	Risico	De beheersing van IT-systemen t.a.v. informatiebeveiliging is niet op orde en daarmee is het systeem, alsmede het algoritme onbetrouwbaar.			
Nr.	**Deelgebied**	**Norm**	**Risico**	**Suggested test steps/evidence**	**Toelichting norm (eventueel)**
GITC.1	Wijzigingenbeheer	Wijzigingen zijn geautoriseerd.	Ongeautoriseerde wijzigingen van het algoritme kunnen plaatsvinden.		
GITC.2		Wijzigingen worden getest.	Ongeautoriseerde wijzigingen van het algoritme kunnen plaatsvinden.		
GITC.3		Wijzigingen worden goedgekeurd met inachtneming van testresultaten.	Ongeautoriseerde wijzigingen van het algoritme kunnen plaatsvinden.		
GITC.4		Functiescheiding bestaat tussen het aanvragen, goedkeuren en doorvoeren van wijzigingen.	Ongeautoriseerde wijzigingen van het algoritme kunnen plaatsvinden.		
GITC.5		Periodieke controle op ongeautoriseerde wijzigingen.	Ongeautoriseerde wijzigingen van het algoritme kunnen plaatsvinden.		
GITC.6	Logische toegangsbeveiliging: Wachtwoordbeheer	Wachtwoorden worden periodiek gewijzigd.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.7		Wachtwoorden zijn van adequate sterkte.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.8		Twee-factor authenticatie wordt gebruikt bij onvertrouwde zones.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.9		Vergrendeling bij inactiviteit.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.10		Wachtwoorden worden alleen versleuteld opgeslagen.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.11		Gebruikersaccounts dienen geblokkeerd te worden na een vooraf ingesteld aantal van vijf tot tien foutieve inlogpogingen.	Wachtwoorden kunnen misbruikt worden om toegang tot het algoritme te krijgen.		
GITC.12	Logische toegangsbeveiliging: Gebruikersbeheer	Gebruikers- en beheerders hebben alleen de toegangsrechten die voor hun functie noodzakelijk zijn.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.13		Gebruikersaccounts en toegangsrechten zijn geautoriseerd.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.14		Functiescheiding bestaat tussen aanvragen, autoriseren en doorvoeren van wijzigingen in gebruikersaccounts en toegangsrechten.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.15		Uitdiensttredingen worden tijdig verwerkt.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.16		Beheeraccounts zijn zo veel mogelijk beperkt en verklaard.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.17		Gebruikersaccounts en beheeraccounts zijn persoonsgebonden.	Onbevoegde personen hebben toegang tot het algoritme.		

GITC.18		Gebruikersaccounts hebben geen directe toegang tot onderliggende componenten.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.19		Gebruikers- en beheeraccounts en toegangsrechten worden periodiek geëvalueerd en de uitkomsten opgevolgd.	Onbevoegde personen hebben toegang tot het algoritme.		
GITC.20	Logische toegangsbeveiliging: Beveiliging van componenten	Er is een actueel inzicht in de applicaties en onderliggende componenten.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.21		Alertering op nieuwe kwetsbaarheden is ingeregeld en systemen worden periodiek gecontroleerd op technische kwetsbaarheden	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.22		Systemen worden tijdig gepatcht en geüpdatet.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.23		Systemen maken geen gebruik van standaard wachtwoorden of backdoors accounts.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.24		Het besturingssysteem draait geen onnodige services.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.25		Het interne netwerk is gescheiden van andere onvertrouwde omgevingen.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.26		Netwerkverkeer en componenten worden actief gemonitord.	Het algoritme is niet goed beveiligd om het risico op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies te voorkomen.		
GITC.27	Back-up & recovery	De periodiciteit van, en het type gegevens op, de back-up sluiten aan bij het belang van het algoritme.	Het algoritme kan bij gegevensverlies niet hersteld worden.		
GITC.28		De back-up gegevens worden op een veilige locatie bewaard waarbij de integriteit van de back-up geborgd blijft.	Het algoritme kan bij gegevensverlies niet hersteld worden.		
GITC.29		Het kunnen terugzetten van de back-up (recovery) wordt periodiek getest.	Het algoritme kan bij gegevensverlies niet hersteld worden.		
Inputkwaliteit					
	Beheersdoelstelling	Er wordt geborgd dat de input dataset representatief is voor de bedoelde populatie en beschermde subpopulaties. Betekenis en herkomst van inputdata zijn eenduidig interpreteerbaar. De inputdata is juist, volledig en consistent. Maatregelen zijn getroffen om vervuiling van inputdata tegen te gaan. Risico's bij hertraining zijn in kaart gebracht.			
	Risico	Op basis van het algoritme worden verkeerde beslissingen genomen door oververtegenwoordiging van een subpopulatie, oneerlijke behandeling van een beschermde subpopulatie, onjuistheid, onduidelijkheid, on- of overvolledigheid, of inconsistente codering van de inputdata, eventueel opzettelijk veroorzaakt door manipulatie van data.			
Nr.	**Deelgebied**	**Norm**	**Risico**	**Suggested test steps/evidence**	**Toelichting norm (eventueel)**

INP.1	Representativiteit van input datasets voor de bedoelde populatie	De doelpopulatie is eenduidig vastgesteld in overeenstemming met het doel van het algoritme.	De inputdata is niet representatief voor de populatie waarover een beslissing genomen wordt, waardoor het algoritme geen goede en/of eerlijke weergave van de werkelijkheid geeft (voor zowel de gehele populatie als deelpopulaties).	In de functionele documentatie wordt de probleemstelling niet geformuleerd in termen van het trainen op een dataset, maar wordt opgebouwd vanuit de te voorspellen doelwaarden met in achtneming van de doelpopulatie.	In een functionele omschrijving is de doelpopulatie vastgesteld in termen van eigenschappen/variabelen van de set waar de doelpopulatie deel van uitmaakt. Bijv. doelpopulatie van tennisballen uit sportartikelen is beschreven in termen van de eigenschappen die de tennisbal kenmerken, nl. kleur, grootte, vorm, textuur, etc. De documentatie bij het algoritme beschrijft duidelijk hoe de doelpopulatie is vastgesteld. De doelpopulatie en diens kenmerken passen bij het doel van het algoritme. Er is beschreven hoe de eigenschappen van de doelpopulatie de doelpopulatie kenmerken. De doelpopulatie is eenduidig vastgesteld in termen van eigenschappen/variabelen van de bron waar de doelpopulatie deel van uitmaakt. De samenstelling van de doelpopulatie is in overeenstemming met het doel van het algoritme. Vaak wordt de doelpopulatie benaderd door beschikbare data uit een herkomstproces dat door het algoritme wordt vervangen, met (doordat impliciet reject inference plaatsvindt) risico op overname en versterking van bestaande biases als gevolg. Stel bijv. dat je voorspelt wie wel of niet een goede werknemer gaat zijn, dan weet je in de regel niets van de mensen die je in het verleden hebt afgewezen; Je baseert je kennis alleen op de mensen die je terecht of onterecht geselecteerd hebt.
INP.2		De input datasets (trainings-, validatie- en testdatasets) zijn, ieder voor zich, voldoende representatieve steekproeven uit de doelpopulatie.	De inputdata is niet representatief voor de populatie waarover een beslissing genomen wordt, waardoor het algoritme geen goede en/of eerlijke weergave van de werkelijkheid geeft (voor zowel de gehele populatie als deelpopulaties).	Samenvattende statistieken input datasets die gebruikt zijn bij de onderbouwing van data en modelkeuze. Exploratieve data-analyse (EDA) naar verschillen in herkomst van de data in de datasets. Ook in de risico-analyse.	De testdataset wordt gezien als de hold-out die na het tunen van de hyperparameters (a.d.h.v. de validatieset) gebruikt kan worden voor een zuivere evaluatie van de performance van het model.
INP.3		Wanneer de populatie anders is dan de doelpopulatie, is er een gestratificeerde, representatieve steekproef getrokken uit de doelpopulatie en wordt deze tenminste gebruikt als evaluatiedataset. Wanneer hier van afgeweken is, is gedocumenteerd waarom de impact van het algoritme en het risico op een gebrek aan representativiteit dit niet vereisen.	De inputdata is niet representatief voor de populatie waarover een beslissing genomen wordt, waardoor het algoritme geen goede en/of eerlijke weergave van de werkelijkheid geeft (voor zowel de gehele populatie als deelpopulaties).	Bij de methode voor het vaststellen van de doelvariabele wordt de beschikbare expertise ingezet om een evaluatiedataset te creëren die als gouden standaard kan functioneren. Door beperkte capaciteit zal deze vaak klein zijn, wat middelen als stratificatie nodig maakt. Stratificatie dekt hier o.a. beschermde groepen. Is afhankelijk van beleidskeuzes tav. ethische risico's. Gebruik de steekproefdata en documentatie van de methode en exploratieve data-analyse als bewijs.	
INP.4		Indien datasets van verschillende herkomsten als input gebruikt worden (dwz. uit verschillende dataverzamelprocessen waar mogelijk andere methodes gebruikt worden), dan is dit verschil in herkomst vastgelegd. De relevante verschillen tussen dataverzamelprocessen zijn beschreven.	De representativiteit van de inputdata wordt niet juist beoordeeld door ontbreken van informatie over de herkomst van de data.	Uit documentatie (data model/glossarium) blijkt dat dataverzamelprocessen zijn beoordeeld op sampling biases. Voorzover deze een rol spelen in de herkomst, wordt dit verschil in herkomst duidelijk vastgelegd in de dataset, zodat herkomst meegewogen kan worden in de dataanalyse en evaluatie.	
INP.5	Representativiteit van input datasets voor beschermde subpopulaties van de bedoelde populatie	De aanwezigheid van beschermde subpopulaties en inputvariabelen die beschermde subpopulaties identificeren (niet zijnde proxy's) is vastgesteld en in overeenstemming met het doel en de impact van het algoritme.	De aanwezigheid van beschermde subpopulaties is niet beoordeeld, of sluit niet aan bij het doel en de impact van het algoritme.	Zie de risico-analyse of DPIA. Vaak worden dit soort attributen weggelaten als inputvariabelen (het zg. non-discrimination principle). Soms heb je ze juist nodig om bias te voorkomen (uitkomstongelijkheid meten, debiasing methodes, oversampling, etc.). De beschermde groepen waarvoor geen oneerlijke uitkomstenongelijkheid mag ontstaan zitten dan evengoed vaak nog in de data. Uitkomstongelijkheid is natuurlijk vooral relevant als er direct of indirect beloningen of straffen worden uitgedeeld. Beschermde status verwijst in ieder geval naar de groepen genoemd in de AVG, aangevuld met ethisch beleid van de eigenaar van het algoritme, maar hangt af van het doel van het algoritme. Bijv. geslacht is vaak geen beschermd kenmerk, maar bijv. in beperkte mate bij personeelsselectie wel.	

INP.6		Indien de impact van het algoritme dit vereist, is data verzameld over beschermde subpopulaties om uitkomstbias te kunnen meten.	Uitkomstbias kan niet gemeten worden door ontbreken van data.	Zie de risico-analyse of PIA. Betreft hier aanvullende evaluatiedatasets die wél inputvariabelen bevatten die beschermde populaties identificeren, die gebruikt kunnen worden om uitkomstbias te meten. Een dergelijke meting toont het bestaan van bias in de dataset aan.	
INP.7		Indien data beschikbaar is die beschermde subpopulaties identificeert, dan wordt vastgesteld of deze informatief is voor de inputvariabelen. De oorsprong van deze informatieve relaties is van een causale interpretatie voorzien in de documentatie, en deze causale interpretatie is in overeenstemming met de risico-analyse.	Het algoritme discrimineert door gebruik van inputvariabelen die indirect beschermde subpopulaties identificeren (proxies).	Met informatief wordt bijv. correlatie, covariatie, relatieve entropie, KL-divergentie bedoeld. Ook wordt wel gesproken van proxies van een beschermde variabele. Volgt uit exploratieve data-analyse. Betreft hier het vermogen om uit een subset van inputvariabelen lidmaatschap van een beschermde te voorspellen. Kan in documentatie zowel in causale termen of in substitutietermen besproken zijn. De gekozen interpretatie zou niet haaks mogen staan op het (ethisch) beleid van de organisatie. Bijv. als geslacht (beschermde groep) met werkervaring (inputvariabele) correleert in een beroepsgroep, dan kan dit verklaart worden door een verschil in de uitstroom van opleidingen in het verleden, of door een personeelsselectiebias. De interpretatie die je geeft beïnvloedt de keuzes die je maakt om al of niet bias te mitigeren. Deze interpretatie dient dus expliciet gemaakt te worden, zodat ook de risico-afweging expliciet is. 1) Bekijk covariaties oid., 2) Gebruik causal diagram, causal loop diagram; Interpretatie volgt uit documentatie.	
INP.8		De input datasets (trainings-, validatie- en testdatasets) zijn voldoende representatief als steekproeven uit de beschermde subpopulaties. Indien de samenstelling van de dataset, de keuze voor een type algoritme, en de impact van het algoritme dit vereisen, worden datasets gebalanceerd voor de beschermde groep. De gekozen strategie is gedocumenteerd en in overeenstemming met de risico-analyse en de gekozen causale interpretatie van informatieve relaties met de inputvariabelen.	De inputdata is niet representatief voor de beschermde subpopulatie waarover een beslissing genomen wordt, waardoor het algoritme geen goede en/of eerlijke weergave van de werkelijkheid geeft.	Betreft hier gedocumenteerde verandering van de samenstelling, bijv. door oversampling, gedurende de preprocessing voor trainen en testen. Kan onderdeel uitmaken van een debiasing-strategie (zie modelkwaliteit). Verdient vaak niet de voorkeur; Liever een in-processing of post-processing oplossing, of reweighting op de input. Oversampling alleen toegestaan als dit uit risico-analyse of PIA volgt.	
INP.9		Er zijn meetbare criteria opgesteld voor toetsing in het in-productiesysteem, doorlopend of periodiek, om de verwachting, zoals deze volgt uit de causale interpretatie, over de ontwikkeling van een informatieve relatie te bevestigen of ontkrachten.	Het algoritme discrimineert door gebruik van inputvariabelen die indirect beschermde subpopulaties identificeren (proxies).	Meetbare criteria volgen uit de functionele documentatie. Met informatief wordt bijv. correlatie, covariatie, relatieve entropie, KL-divergentie bedoeld. Ook wordt wel gesproken van proxies van een beschermde variabele. Gedocumenteerd en geïmplementeerd. Idee is hier dat vastgestelde proxies van beschermde groepen ook over de tijd gevolgd kunnen worden, zodat bijv. geconstateerd kan worden dat verandering optreedt. Ideaal is doorlopende monitoring, maar ook een periodieke evaluatie is mogelijk.	Bijv: door nudging neemt de informatieve relatie tussen postcode en politietoezicht toe.
INP.10		Indien de inputdata data bevat die informatief is voor lidmaatschap van een beschermde subpopulatie of voor persoonsgegevens, dan wordt geborgd dat deze aan dezelfde eisen voldoet als inputdata die onder de AVG valt.	De verwerking van de inputdata voldoet niet aan de wettelijke verplichting AVG.	Zie de DPIA. Stel vast dat toepassing van AVG normen op de inputdata procesmatig geborgd is, en dat ook als uit exploratieve data-analyse gebleken is dat bepaalde data als proxy van een beschermde groep of persoonsgegeven kan functioneren, deze proxies betrokken zijn.	
INP.11	Begrijpelijkheid van inputvariabelen	De keuze voor de gebruikte input variabelen is onderbouwd en in overeenstemming met het doel van het algoritme.	De prestatie van het algoritme is lager door verkeerde keuze van inputvariabelen.	Onderbouwing gebruikte variabelen in documentatie. Wanneer zijn ze relevant, inclusief historie van exploratieve data-analyse (EDA). Ook de robuustheid van de causale relaties tussen input data en de te voorspellen doelvariable kan een belangrijke rol spelen, afhankelijk van doel algoritme.	

INP.12		Wanneer de inputvariabelen aangepast worden (pre-processing), zijn de aanpassingen onderbouwd en in overeenstemming met het doel van het algoritme.	De prestatie van het algoritme is lager door onjuiste aanpassing van de inputdata.	In de documentatie is beschreven welke pre-processing stappen zijn uitgevoerd en waarom dit is gedaan. De effecten van de verschillende stappen op de data en het model zijn beschreven.	Voor beeldherkenning bestaan de inputvariabelen uit de foto's en annotatie. Daarnaast is het mogelijk dat er extra gegevens over de foto worden gelegd, als bijv. grenzen van objecten op basis van contrast. Ook is het mogelijk om de input foto's aan te passen door bijvoorbeeld de rgb pixels aan te passen of een specifiek kleurkanaal te gebruiken. Augmentatie van beelden valt ook onder pre-processing.
INP.13		De gebruikers van het algoritme en andere belanghebbenden zijn betrokken bij de keuze van de inputvariabelen tijdens de ontwikkeling van het algoritme, en in de productiefase is de gebruikers bekend welke inputvariabelen gebruikt worden voor de voorspelling van de doelvariabele(n).	De ontwikkelaar van het algoritme kiest verkeerde inputvariabelen door gebrek aan kennis over het doel en de context van het algoritme.	Gebruikers met veel ervaring in het doen van voorspellingen kunnen de voorspellingen die het algoritme doet beter op waarde schatten als ze begrijpen welke input gebruikt wordt. Zijn gebruikers beschikbaar voor interview die een rol hebben gehad bij de ontwikkeling? Kunnen zij bevestigen dat in de in-productie context informatie beschikbaar is over welke inputdata gebruikt wordt.	
INP.14		Het is vastgesteld dat de labels van de inputdata eenduidig en in overeenstemming met het doel van het algoritme zijn.	De prestatie van het algoritme is lager door onjuiste labeling van de inputdata.	Documentatie/interviews/metadata	Niet altijd maakt de organisatie de labels zelf. Er moet beschreven zijn waarom deze labels gebruikt kunnen worden voor het doel van het algoritme. Wanneer de organisatie zelf labelt zijn er vooraf duidelijke afspraken gemaakt over wat de labels betekenen, zodat er op dezelfde manier door verschillende personen gelabeld kan worden.
INP.15		De inputvariabelen zijn begrijpelijk en eenduidig: alle variabelen zijn duidelijk beschreven, zowel in termen van eigenschappen van de doelpopulatie als in termen van herkomst (dataverzamelingsproces).	Beslissingen worden genomen op basis van gegevens die verkeerd begrepen zijn, of omdat het algoritme iets anders doet dan verwacht omdat inputvariabelen een andere betekenis hebben dan verwacht.	Glossarium/naamconventielijst/data dictionary/metadata/data model. Belangrijk is hier ook dat duidelijk te reconstrueren wanneer en met welk doel data ontstaan is.	Bijv. pixels die 20m bij 20m beschrijven passen niet bij het doel van het herkennen van personen.
INP.16	Juistheid en volledigheid	Datasets voor (her)training, testen en evaluatie zijn voldoende recent, voorzover de veranderlijkheid van de doelpopulatie en de impact en het doel van het algoritme dit vereisen.	De prestatie van het algoritme is lager door ouderdom van de gebruikte inputdata.	Betreft met name de vertraging tussen de periode die de datasets waatrop getraind, getest, en geëvalueerd wordt, en het tijdvak waarin het algoritme in productie is. Hierdoor neemt het risico toe dat het algoritme niet generaliseerbaar blijkt naar de in-productie context, door omgevingsveranderingen. Wat tijdig is is afhankelijk van de verwachte veranderlijkheid van de relatie tussen de inputdata en de te voorspellen doelvariabele. Eventueel is dit besproken met een domein expert.	
INP.17		De inputdata in productie is tijdig, volledig en definitief beschikbaar voor het doel van het algoritme	Het algoritme kan in productie niet gebruikt worden door het ontbreken van inputvariabelen die in de ontwikkelfase in de inputdata wel al beschikbaar waren en gebruikt zijn.	De ontwikkelaars bouwen een tijdmachine: data uit de toekomst wordt gebruikt om de toekomst te voorspellen. Dit wordt wel het onderscheid tussen een steekdatum en een zichtdatum genoemd: wat we nu (zichtdatum) over een gebeurtenis (steekdatum) weten, is niet wat we op de steekdatum wisten over de gebeurtenis. Bijv. het feitenrelaas uit een vonnis kan – met veel succes – gebruikt worden om de inhoud van het vonnis te voorspellen. Probleem is natuurlijk dat dit feitenrelaas is opgeschreven op een moment (de zichtdatum) dat het hof al wist naar welk vonnis ze toe aan het argumenteren waren, en het feitenrelaas daar overduidelijk sporen van bevat, omdat de gekozen relevante feiten (die nover de steekdatum gaan) degenen zijn die nodig zijn om tot dat specifieke vonnis te komen. Als inputdata gewijzigd wordt naar aanleiding van de vaststelling van de doelvariabele (tussen steek- en zichtdatum), dan is de trainings- en testset (op de zichtdatum) geen accurate weergave van welke inputdata beschikbaar was op het moment (de steekdatum) dat het nuttig was om de doelvariabele te kunnen voorspellen. Daardoor zal het algoritme in productie niet werken (als inputdata ontbreekt), of misleidend werken (als inputdata in de toekomst veranderd wordt, en dit onopgemerkt blijft door de ontwikkelaars).	

INP.18		De inputdata is juist en volledig ingelezen in de data-analysetool en sluit aan met de bron.	Het algoritme geeft verkeerde uitkomsten doordat de inputdata niet volledig, niet juist of inconsistent is.	Onderbouwing aansluiting input dataset met bron van herkomst, eventueel te toetsen door reproductie. Onderbouwing door technische documentatie, alleen in geval van ernstige twijfel reproductie van een steekproef.	
INP.19		De herkomst van de inputdata uit dataverzamelingsprocessen is herleidbaar en blijvend reproduceerbaar vastgelegd.	De inputdata en daarmee ook de output van het algoritme is onvolledig of onbetrouwbaar.	Overzicht databronnen; metadata over herkomst uit een dataverzamelingsproces; principieel bestaat de mogelijkheid de datasets te reproduceren uit de bron. Als kwaliteitsverhogende stappen zijn uitgevoerd, zijn deze te reproduceren.	
INP.20		De input dataset uit de bron (het dataverzamelingsproces) is verifieerbaar volledig en betrouwbaar genoeg voor het doel van de data-analyse.	De inputdata en daarmee ook de output van het algoritme is onvolledig of onbetrouwbaar.	De dataverzamelingsprocessen die als bron dienen zijn gedocumenteerd. Data quality is onderzocht en stappen zijn genomen om evidente vervuiling op te lossen. Datapreparatie-onderbouwing en code voor preprocessing. Preprocessing is reproduceerbaar en door middel van sampling en analyse getest. Eventueel reproductie van een sample van data uit de dataset uit de bron. Vastgesteld wordt zowel óf de aanpak effectief is, en of de sample correct gereproduceerd wordt.	Hoe komt de data tot stand bij de bron?
INP.21		De betrouwbaarheid van de bron is aangetoond.	De inputdata en daarmee ook de output van het algoritme is onvolledig of onbetrouwbaar.	Onderbouwing bronkeuze, belangenafweging bij het vertrouwen van de bronhouder, en chain of custody. Aanwezigheid overeenkomsten controleren, algemene kenmerken herkomstproces beschreven in documentatie.	
INP.22	Consistente codering van de inputdata	De inputdata is consistent gecodeerd, waarbij dezelfde bronnen, coderingen en verwijzingen gebruikt worden (op veldniveau).	De inputdata bevat inconsistenties.	Zie data model/glossarium etc. Onderzoek en preprocessing stappen voor data quality, oog voor subtiele verschillen in functie van ongeveer dezelfde informatie (bijv. adres, dat kan verschillen afhankelijk van toepassing: bezoek, pakket, briefpost, etc., tijdelijk vs. permanent, etc.)	
INP.23		Maatregelen zijn getroffen in dataverzamelingsprocessen om de consistente codering van inputdata te waarborgen.	De inputdata bevat inconsistenties.	Documentatie: gekeken is naar oorzaken van mogelijke inconsistenties in codering; Beheersmaatregelen zijn genomen bij de bronhouder indien nodig en mogelijk. Aanwezigheid overeenkomsten controleren, algemene kenmerken herkomstproces beschreven in documentatie.	
INP.24		Elke entiteit uit de werkelijkheid komt slechts als één record voor in de inputdata, tenzij nadrukkelijk anders bedoeld.	De inputdata geeft de werkelijkheid niet goed weer omdat er dubbelingen in voorkomen.	Zie documentatie preprocessing voor data quality en risico-analyse data poisoning. Bedoeld wordt een bijectieve relatie tussen domain-entiteit en data-entiteit, en normalisatie van sleutels. Zie technische documentatie, eigenschappen algoritme of tool, of code of steekproef data	Voorbeelden waarbij meer dan 1 record in de inputdata gebruikt worden zijn oversampling/augmentatie/beelden onder verschillende hoeken verstaan. De organisatie moet deze keuze beargumenteren.
INP.25	Bescherming	Het risico van datavervuiling in de herkomstprocessen is onderzocht.	De inpudata is vervuild.	Het risico op vervuilde data uit de dataverzamelingsprocessen is onderzocht in de risico-analyse, rekening houdend met het doel van het algoritme.	Datavervuiling kan onbewust of bewust plaatsvinden. Daarnaast kan datavervuiling bij de inputdata zelf, of in de annotatie van de data voorkomen.
INP.26		De chain of custody van dataverzamelingsproces naar input-dataset is gesloten.	De inpudata is vervuild.	Alle preprocessing-bewerkingen zijn reproduceerbaar en worden steekproefsgewijze gereproduceerd. Bevoegdheid en mogelijkheid om data te bewerken is gelimiteerd. Aanwezigheid overeenkomsten controleren, algemene kenmerken herkomstproces beschreven in documentatie.	
INP.27		In productie wordt de samenstelling van de inputdata getoetst op verandering ten opzichte van de tijdens trainen, testen en evalueren geobserveerde samenstelling.	De prestatie van het algoritme is lager doordat de samenstelling van de inputdata in productie is veranderd t.o.v. de ontwikkelfase.	De verwachtingen die we hebben over de input data in productie zijn geöperationaliseerd als meetbare criteria en geïmplementeerd in het systeem in productie, of worden door periodieke toetsing (evt. op steekproeven) gerealiseerd. Gebaseerd op de risico-analyse kan tijdigheid van toetsing een heel belangrijke rol spelen. Denk bijv. aan de Bulgarenfraude bij BD Toeslagen; Hoewel de locatie vanwaar een aanvraag ingediend wordt an sich niet relevant hoeft te zijn voor het risicoprofiel van een individuele aanvraag, had het ontstaan van een piek in aanvragen vanuit een bepaalde locatie bellen moeten laten rinkelen. Dit had door middel van	

				doorlopende monitoring door het systeem zelf gerealiseerd kunnen worden. Deze monitoring moet gevoelig genoeg zijn om snel genoeg een afwijking te detecteren. Zie functionele documentatie en interpretatie als meetbare criteria, documentatie EDA op in-productie-data Blijven gemiddelden, varianties, etc. gelijk?	
INP.28		Het risico op data poisoning is getest in het geval van hertraining van het model.	Het algoritme is niet veilig tegen datapoisoning en corruptie.	Alleen van toepassing als hertraining van het model nodig is. De risico-analyse stelt vast of er risico op data poisoning in deze fase bestaat, en indien nodig zijn beheersmaatregelen genomen.	
INP.29	Hertraining	In het geval dat hertraining plaats zal vinden is getest of de inputdata nog representatief is en niet voor een bias zorgt.	Nudging vindt plaats, zodanig dat biases richting rejects bij hertraining versterkt worden.	Betreft hier een risico op nudging zodanig dat biases richting rejects bij hertraining versterkt worden. Indien alleen rejects in behandeling worden genomen en de resultaten hiervan worden gebruikt voor hertraining, moet deze data aangevuld worden met een random steekproef.	
Modelkwaliteit					
	Beheersdoelstelling	Maatregelen borgen dat het model een zo goed en eerlijk mogelijke weergave van de werkelijkheid is. Het model is eerlijk ten aanzien van beschermde subpopulaties en individuen. En het model is controleerbaar.			
	Risico	Het model is geen goede weergave van de werkelijkheid door manipulatie, onvoldoende beoordeling van de nauwkeurigheid en robuustheid, of doordat wijzigingen onzorgvuldig plaatsvinden, het model is niet eerlijk, of het functioneren van het model op voorgaande punten is niet controleerbaar.			
Risico	**Deelgebied**	**Norm**	**Risico**	**Suggested test steps/evidence**	**Toelichting norm (eventueel)**
MOD.1	Integriteit	De keuze voor het modeltype en onderliggende (hyper)parameters is onderbouwd en sluit aan bij het beoogde doel van het algoritme.	Het model kan gemanipuleerd zijn/worden om gewenste uitkomsten te forceren.	Technisch en functioneel ontwerp/omschrijving, inclusief de keuzes die gemaakt zijn. Denk hierbij aan modelkeuze, hyperparameterkeuze e.d.	Voor een neuraal netwerk is het tevens belangrijk dat het aantal nodes past bij de inputdata. Er kan bijvoorbeeld data verloren gaan wanneer er minder input nodes dan pixels zijn.
MOD.2		Het model wordt alleen gebruikt voor het doel waarvoor het ontwikkeld is, en geëvalueerd met prestatiecriteria die volgen uit het doel van het algoritme.	Het model wordt voor andere doeleinden gebruikt dan oorspronkelijk bedoeld en niet aan de juiste prestatiecriteria getoetst.	Samenhang risico-analyse en functionele documentatie met het daadwerkelijk gebruik van het algortime in produktie en de performance-indicatoren die daar gebruikt worden. Deze norm sluit ook gebruik voor secondaire doelen waarvoor het handig blijkt uit.	
MOD.3		De keuze voor modeltype en onderliggende (hyper)parameters is getoetst door middel van peer review.	De keuze voor modeltype en (hyper)parameters is niet optimaal door fouten of ontbrekende kennis van de ontwikkelaar van het algoritme.	4 ogen principe is toegepast. Vraag om een verslag van een peer review.	
MOD.4	Doelmatigheid	De acceptatiecriteria voor het model op het gebied van robuustheid, sensitiviteit en nauwkeurigheid zijn operationaliseerbaar, duidelijk gedocumenteerd en volgen uit de functionele eisen en een risicoafweging van beoogde gebruikers van de voorspellingen van het model en andere belanghebbenden die een duidelijk belang hebben bij de criteria.	Het is niet duidelijk aan welke acceptatiecriteria het model moet voldoen, of de gestelde criteria passen niet bij de functionele eisen en risicoafweging.	Toestemming voor het inzetten van het algoritme volgt uit een duidelijke afspraak over wat de acceptabele risico's op het gebied van betrouwbaarheid en robuustheid etc. zijn in de beslissingscontext waarin het algoritme ingezet wordt.	

MOD.5	Nauwkeurigheid	De wiskundige aannames die bij het modeltype horen zijn in overeenstemming met de statistische eigenschappen van de data.	De wiskundige aannames die bij het gekozen modeltype horen passen niet bij de statistische eigenschappen van de data.	Vastgesteld in technisch en functioneel ontwerp/omschrijving, inclusief de keuzes die gemaakt zijn. Vergelijken met samenvattende statistieken input datasets, en eventuele preprocessing om te herschalen.	
MOD.6		Op basis van de functionele eisen en risico-afwegingen zijn acceptatiecriteria vastgesteld die operationaliseerbaar zijn als meetbare nauwkeurigheidsmaten op de voorspellingen die gedaan worden door het model. Deze maat is passend voor het type voorspelling en de risico's van de toepassing van deze voorspelling in een daaropvolgende beslissing.	Het model voldoet niet aan de (juiste) acceptatiecriteria.	Beschrijving algoritme of projectbeschrijving stelt een duidelijke definition of success vast, en deze is vertaalbaar naar nauwkeurigheid en betrouwbaarheid. Accuracy wordt gemeten op een manier die recht doet aan de doelvariabele (ja/nee beslissing, ranking, kans etc.), en aan de op basis van risico geformuleerde wensen op het gebied van bijv. FPR en FNR.	Opmerking: Mogelijk dubbel na regel 37, misschien goed om per deelgebied nog wel extra specificering zoals in deze norm te geven. Kan als subnorm bij deelgebied doelmatigheid.
MOD.7		De nauwkeurigheid van het model is aantoonbaar berekend met behulp van een hold-out dataset en getoetst aan de acceptatiecriteria. De keuze voor de hierbij gebruikte rekenmethode is onderbouwd, en de test-dataset is aantoonbaar niet betrokken geweest bij de keuze van een modeltype en parameters (holdout dataset), tenzij dit gezien het doel en de impact van het algoritme niet nodig is.	De nauwkeurigheid van het model is niet goed beoordeeld waardoor verkeerde outputdata wordt gegenereerd.	Onderbouwing acceptatie geschatte parameters, inclusief de door het model geschatte parameters/diagnostische gegevens (denk aan: coefficienten, significantie, betrouwbaarheidsintervallen). 1) Controleer of een definition of success is vastgesteld die operationaliseerbaar is als nauwkeurigheidsgrens, en 2) of deze passend getoetst is op het model.	De test-data is aantoonbaar out-of-sample; dat wil zeggen op geen enkel moment bij het trainen betrokken geweest. Ook wanneer het model aangepast wordt. Er is gebruik gemaakt voor cross-validatie om een bias door een toevallige splitsing tussen een train-, validatie- en hold-out set te voorkomen of er is beschreven waarom het niet nodig was een dergelijke methode te gebruiken.
MOD.8		De betrouwbaarheid van het model is aantoonbaar getoetst aan de acceptatiecriteria voordat het model in productie wordt genomen.	De betrouwbaarheid van het model is niet goed beoordeeld waardoor verkeerde outputdata wordt gegenereerd.	Meest voor de hand liggende manier om betrouwbaarheid te meten is het betrouwbaarheidsinterval op de doelvariabele. Bij een neuraal netwerk is het niet mogelijk om een betrouwbaarheidsinterval op te stellen. Het is wel mogelijk de variatie van de resultaten bij cross-validatie te bekijken als maat van betrouwbaarheid/robuustheid van het model.	
MOD.9		Getoetst aan de acceptatiecriteria presteert het model gelijkwaardig op nieuwe input data en op de trainings- en testdatasets voor het doel en de impact van het algoritme. Variatie is gedocumenteerd en wordt verklaard.	Het model is te specifiek gevormd naar de trainingset, waardoor het slechter presteert met nieuwe data dan met de trainingsset.	Train en testresultaten, inclusief onderbouwing waarom deze afdoende zijn (bijv. geen overfitting). Omdat meestal alleen voor een geselecteerde groep een uiteindelijke vaststelling van de doelvariabele plaatsvindt (de rejects worden buiten beschouwing gelaten) is de nieuwe productiedata vaak niet representatief voor de doelpopulatie. Variatie is dus vaak verklaarbaar.	
MOD.10		De prestaties van het model op nieuwe inputdata worden in productie doorlopend of periodiek getoetst en vergeleken met de vooraf vastgestelde acceptatiecriteria. De toetsing is voldoende tijdig voor het doel en de impact van het algoritme. Als de acceptatiecriteria niet meer gehaald wordt, wordt het algoritme uit productie genomen of worden tijdig andere remediërende maatregelen genomen.	De prestatie van het model voldoet niet meer aan de acceptatiecriteria door verandering van omstandigheden over de tijd heen.	Aanwezigheid monitoring en audit trail als het systeem in productie is. Definition of success wordt blijvend getoetst. Proces van evaluatie is aantoonbaar geborgd.	
MOD.11	Robuustheid	Het model is getest op sensitiviteit: de range van de te verwachten outputdata is in kaart gebracht en getoetst aan acceptatiecriteria voor betrouwbaarheid en robuustheid.	Het model is instabiel omdat het gevoelig is voor kleine veranderingen. Het model is niet getest op sensitiviteit van de parameters of invloed van variabelen, waardoor worst case scenarios onbekend zijn.	Resultaten sensitiviteitsanalyse, met onderbouwing acceptatie (tolerantiegrenzen) Op skills niveau: Is het team per direct in staat hypothetische cases door het systeem te voeren?	Sensitiviteit is test evidence, rubuustheid is de norm

MOD.12	Herevaluatie	De nauwkeurigheid, betrouwbaarheid, en de robuustheid van het model worden geherevalueerd na iedere wijziging aan het model in productie die deze kunnen beïnvloeden.	Het model voldoet na een wijziging niet meer aan de acceptatiecriteria.	Overzicht data- en/of modelwijzigingen. Zie functionele en technische documentatie, documentatie QA processen.	
MOD.13		Als sprake is van uitsluiting van een deel van de populatie zodanig dat de doelvariabele(n) voor deze groep nooit vastgesteld worden (zg. rejects) dan worden de risico's van gebruik van evaluatiedata uit productie gedocumenteerd en de acceptatiecriteria aangepast voor herevaluatie.	Nudging vindt plaats, zodanig dat biases richting rejects bij hertraining versterkt worden.	Zie risico-analyse en acceptatiecriteria specifiek vastgesteld voor een herevaluatie. Dit risico zou in de risico-analyse al gedocumenteerd moeten zijn, en bekend bij de ontwikkelaars. Rejects zijn een reden om variatie in prestaties van het systeem, met uitleg, te accepteren.	In je risicoanalyse moet je bij voorbaat bedenken wat het effect van 'nudging' kan zijn: wat gaat de cyclus van het trainen op voorgaande jaren doen met de accuraatheid met in het hoofd houdende van wat je weet over de stabiliteit van de populatie en het probleem.
MOD.14	Eerlijkheid	De acceptatiecriteria voor het model op het gebied van eerlijkheid ten aanzien van beschermde groepen en individuen zijn operationaliseerbaar, duidelijk gedocumenteerd en volgen uit de functionele eisen en een risicoafweging van beoogde gebruikers van de voorspellingen van het model en andere belanghebbenden die een duidelijk belang hebben bij de criteria.	Het model voldoet niet aan acceptatiecriteria t.a.v. beschermde subpopulaties.	Controleer of in de definition of success criteria zijn geformuleerd voor eerlijkheid en deze vertaalbaar zijn naar een meetbaar criterium. Toestemming voor het inzetten van het algoritme volgt uit duidelijke acceptatiecriteria over hoe nauwkeurig het algoritme voorspelt voor beschermde groepen, en wat de acceptabele risico's op het gebied van privacy, discriminatie van beschermde groepen, uitsluiting van individuën, etc. zijn in de beslissingscontext waarin het algoritme ingezet wordt. Voor discriminatie hangt dit samen met de causale interpretatie die aan informatieve relaties tussen lidmaatschap van beschermde groepen en inputvariabelen: maten voor discriminatie zijn gebaseerd op vooronderstellingen over de ground truth omtrent dit soort relaties (bijv. predictive equality veronderstelt dat subpopulaties gelijk zijn, en in de data observeerbare ongelijkheden die tot uitkomstenongelijkheid leiden zijn dan gevolg van bias die weggewerkt moet worden, en predictive parity maten eisen alleen gelijke nauwkeurigheid voor de subpopulaties).	
MOD.15	Reproduceerbaarheid	Iedere model is reproduceerbaar aan de hand van inrichtingsdocumentatie en de gebruikte trainingsdataset. Variatie tussen het productiemodel en het gereproduceerde model is vermeden, of anders verklaarbaar uit de eigenschappen van het gebruikte modeltype.	De totstandkoming van het model is onherleidbaar, waardoor transparantie en verantwoording niet mogelijk is.	Overige inrichtingsdocumentatie. De modelparameters dienen zodanig gedocumenteerd te worden (inclusief seed voor evt. randomizers) dat variaties te vermijden zijn, tenzij dit door de gebruikte implementatie van een modeltype niet mogelijk is. Bijv. een aantal implementaties van modeltypes die gebruik maken van de GPU bij het trainen zijn non-deterministisch afhankelijk van de gebruikte hardware, dwz. dat een andere machine een iets ander model zal creëren. Interview en documentatie. Bij twijfel om reproductie vragen.	

Outputkwaliteit		
	Beheersdoelstelling	Geborgd is dat beslissingen worden genomen op basis van juiste outputdata afkomstig uit het algoritme. Er ontstaat geen uitkomstenongelijkheid voor beschermde subpopulaties of door onbedoelde identificeerbaarheid van individuen. De uitkomsten uit het model zijn reproduceerbaar.
	Risico	Op basis van het algoritme worden verkeerde beslissingen genomen door onjuistheid, onduidelijkheid, on- of overvolledigheid van de outputdata, of de outputdata is niet tijdig genoeg om mee te wegen in de beslissing, de outputdata nodigt uit tot oneerlijke beslissingen of onbedoeld gebruik voor andere doelen, of de audit trail is onvoldoende compleet, betrouwbaar of tijdig om vast te kunnen stellen of verkeerde beslissingen genomen kunnen worden.

Risico	Deelgebied	Norm	Risico	Suggested test steps/evidence	Toelichting norm (eventueel)

OUT.1	Juistheid	De outputdata van het algoritme worden periodiek (steekproefsgewijs) of doorlopend getoetst op juistheid en nauwkeurigheid, aan de hand van meetbare criteria vastgesteld als uitwerking van eisen die als onderdeel van het functioneel ontwerp zijn vastgelegd.	Fouten of afwijkende outputdata van het algoritme worden niet gesignaleerd, doordat hier geen aandacht voor is of doordat niet duidelijk is aan welke eisen de outputdata moet voldoen.	Zoek periodieke rapportage van nauwkeurigheid en juistheid, controleer of deze coherent volgen uit de functionele documentatie.	
OUT.2		Outputdata die de meetbare prestatiecriteria overschrijdt wordt tijdig gedetecteerd en geanalyseerd, en aan afwijkingen wordt opvolging gegeven om eventuele fouten te herstellen en/of voorkomen.	Fouten of afwijkende outputdata van het algoritme worden niet of te laat gesignaleerd, of aan geconstateerde afwijkingen wordt geen opvolging gegeven.	Voorbeeld melding en verwerking afwijking(en). Vraag of er ooit iets mis gaat, ga dan follow up na.	
OUT.3		De middelen voor het tijdig analyseren en corrigeren van afwijkingen van de meetbare prestatiecriteria zijn voldoende geborgd.	Ondanks het bewustzijn van fouten of afwijkende outputdata van het algoritme wordt hier geen opvolging aan gegeven door het ontbreken van de benodigde middelen.	Tijdige inzetbaarheid van ontwikkelaars is geborgd, behoud van ervaring en deskundigheid is geborgd. Borging door middel van een proces en beschikbaarheid budget, of door overeenkomsten. Vraag naar ervaringen ontwikkelaars met follow-up.	
OUT.4	Begrijpelijkheid	De outputvariabelen zijn begrijpelijk en eenduidig interpreteerbaar: alle variabelen zijn duidelijk beschreven, in termen van eigenschappen van de doelpopulatie en in termen van herkomst (het algoritme).	Beslissingen worden genomen op basis van outputdata die verkeerd begrepen zijn, omdat outputvariabelen een andere betekenis hebben dan verwacht.	Glossarium/naamconventielijst/data dictionary/metadata/data model Kies een variabele, zoek een omschrijving in de documentatie en modellen.	
OUT.5		De gebruikers van het algoritme zijn aan de hand van de outputvariabele(n) in staat aan andere belangrijke belanghebbenden uit te leggen wat de voorspelling door het algoritme betekent en hoe deze voorspelling tot stand is gekomen.	Gebruikers van het algoritme geven een onjuiste uitleg over de uitkomsten van het algoritme aan andere belanghebbenden, omdat outputvariabelen een andere betekenis hebben dan verwacht.	Dit is een algemene uitlegbaarheidsnorm. Denk hierbij ook aan goed begrip van de limitaties en onzekerheid van de uitkomsten van het algoritme. Interview een gebruiker.	
OUT.6	Volledigheid	Indien de impact van het algoritme dit vereist, wordt de volledige outputdata mét de inputdata bewaard.	Uitkomsten uit het algoritme ontbreken door onvolledigheid van de outputdata.	Aansluiting output met de input door middel van metadata. Dit is een voorwaarde om reproduceerbaarheid te testen. Glossarium/naamconventielijst/data dictionary/metadata/data model	
OUT.7		Elke legitieme uitkomst van toepassing van het algoritme komt voor in de outputdata, slechts als één record, tenzij nadrukkelijk anders bedoeld, en in de outputdata komen alleen legitieme uitkomsten gebaseerd op legitieme inputdata voor.	De outputdata geeft de werkelijkheid niet goed weer omdat er dubbelingen in voorkomen.	Zie documentatie data quality en risico-analyse data poisoning. Bedoeld wordt dat er een bijectieve relatie tussen inputdata, uitkomst en outputdata-entiteit is, zodat schadelijke side-effects vermeden worden. Bijv. een toeslag wordt twee keer uitbetaald, omdat deze tweemaal door het algoritme gehaald is, of tests verricht door ontwikkelaars beïnvloeden de outputdataset. inclusief het vereiste van unieke sleutels voor database entries, zoals uit werkproces, de code of de dataset zelf vast te stellen is.	
OUT.8	Tijdigheid	De outputdata is tijdig beschikbaar voor het doel van het algoritme.	Vertraging in beschikbaarheid van de output kan impact hebben op personen, de maatschappij of de bedrijfsvoering.	Documentatie afwerking output, eventueel in de functionele documentatie van het algoritme zelf: voorbeeld waarin beschreven is hoe de output leidt tot beslissingen. Feedback vanuit de gebruikscontext in productie betreffende tijdigheid. Hierbij dient specifiek gelet te worden op hertraining van het model, en de consequenties voor tijdigheid. Daarnaast moet ook de mogelijkheid van tijdmachine-effecten (zie tijdigheid beschikbaar inputvariabelen) getoetst worden.	
OUT.9	Eerlijkheid	De uitkomsten van het model worden doorlopend of periodiek getoetst op uitkomstenongelijkheid voor beschermde subpopulaties als data beschikbaar is over het lidmaatschap van beschermde subpopulaties. Afhankelijk van de	De outputdata bevat ongewenste uitkomstenongelijkheid voor beschermde subpopulaties.	Een audit trail van gemeten uitkomstenongelijkheid dient beschikbaar te zijn als de data om dit te toetsen beschikaar is en het systeem in productie is. Let ook op	

		acceptatiecriteria voor eerlijkheid, en impact en het doel van het algoritme worden remediërende maatregelen getroffen.		dat uitkomstenongelijkheid met een juiste maat getoetst wordt. Proces van evaluatie is aantoonbaar geborgd.	
OUT.10		Indien de outputdata, inclusief eventuele subsets die bedoeld zijn als onderdeel van uitlegbaarheid van de beslissing, data bevat die informatief is voor lidmaatschap van een beschermde subpopulatie of informatief is voor persoonsgegevens, dan wordt geborgd dat deze aan dezelfde eisen voldoet als inputdata die onder de AVG valt.	De verwerking van de outputdata voldoet niet aan de wettelijke verplichting AVG.	Stel vast dat toepassing van AVG normen op de outputdata procesmatig geborgd is. Omdat hier oneerlijke behandeling van individuën aan de orde is, kan borging van eerlijke behandeling alleen tot stand komen door normen voor omgang met persoonsgegevens. De oneerlijke behandeling is niet kwantificeerbaar op geäggregeerd niveau. Informativiteit zou moeten volgen uit een EDA, als een dataset over de beschermde populatie beschikbaar is. Dit is alleen relevant als de output voldoende complex is om een EDA te rechtvaardigen. Iedere subset die als sterk informatief beoordeeld wordt zou als potentieel beschermd kenmerk besproken moeten worden in de PIA.	
OUT.11	Reproduceerbaarheid	Iedere individuele output is in het kader van een bezwaar- of beroepsprocedure, of een andere klachtprocedure met vergelijkbare strekking, met behulp van hetzelfde model en dezelfde inputdata te reproduceren, voor zover bewaartermijnen voor data dit mogelijk maken.	Bij een bezwaar- of beroepsprocedure is niet na te gaan waarom het algoritme voor een specifiek individu tot een bepaalde uitkomst is gekomen.	Eerlijke besluitvorming vereist dat de beslissing reproduceerbaar is. Het kan voorkomen dat na een bepaalde termijn de data niet meer bewaard wordt; Deze termijn zou moeten samenvallen met de bewaartermijn.	
OUT.12		De outputdata is op hoofdlijnen reproduceerbaar aan de hand van de gebruikte inputdata, het modeltype en (hyper)parameters. De variatie tussen de oorspronkelijke en de gereproduceerde outputdata is verklaarbaar vanuit de eigenschappen van de gebruikte technologie.	De outputdata van het algoritme is onherleidbaar, waardoor transparantie en verantwoording niet mogelijk is.	Overzicht databronnen; metadata over herkomst uit een dataverzamelingsproces; principieel bestaat de mogelijkheid de datasets te reproduceren uit de bron. Steekproefsgewijs kan een reproductie plaatsvinden als twijfel is aan reproduceerbaarheid. Machine learning technieken zijn over het algemeen niet geheel deterministisch, waardoor enige variatie zal ontstaan. Als de inputdata voldoende geborgd is, het model reproduceerbaar wordt geacht, en de outputdata voldoende geborgd is, is dit aannemelijk.	