1 Projectomschrijving

2 Scope van deze technische review

3 Resultaten van de technische review

4 Conclusie & aanbevelingen

Rapportage technische review blurringsalgoritme gemeente Amsterdam

juni 2023

De gemeente Amsterdam wil op verantwoorde wijze technologie inzetten om haar dienstverlening te verbeteren. Eén van de potentiële technologische ontwikkelingen hiervoor is het inzetten van een algoritme om persoonsgegevens op foto's te detecteren en onzichtbaar te maken, ook wel blurren genoemd. Door gebruik te maken van dit algoritme kunnen panoramabeelden van de buitenruimte in de toekomst hergebruikt worden, zonder dat hiermee persoonsgegevens verder worden gedeeld. Omdat de gemeente alleen algoritmes wenst in te zetten die volgens de laatste ethische en privacystandaarden werken, is besloten om zelf een blur-algoritme te ontwikkelen die aan deze standaarden voldoet.

Verdonck, Klooster & Associates (VKA) is gevraagd om een technische review uit te voeren op het ontwikkelde algoritme, waarbij de focus ligt op de technische vertaling van de ethische en privacystandaarden tijdens het ontwikkelproces. Met de vraag naar deze review, laat de gemeente zien dat zij het verantwoord inzetten van algoritmes belangrijk vindt en hier mee bezig is en blijft.

Omdat de review een momentopname (april 2023) is, is op een tweede moment (juni 2023) een deel van de documenten die nog niet gereed of opgesteld waren ten tijde van de initiële review, heeft gemeente Amsterdam gevraagd om deze mee te nemen in een herbeoordeling.

In deze technische review heeft VKA aan de hand van literatuurstudie, interviews en code review gekeken naar de sluitendheid, de volledigheid en het proces van de ontwikkeling van het algoritme. Binnen de scope van deze review zijn team-leden van het Computer Vision Team (hierna: CVT) - bestaande uit twee ontwikkelaars, tech lead, business analist en de projectmanager) geïnterviewd. Dit rapport geeft inzichten in de resultaten van de technische review zowel in initiele review als de herbeoordeling.

Deze rapportage is als volgt opgebouwd. Allereerst geven we een korte omschrijving van het blurringsalgoritme en hoe de review is uitgevoerd. Daarna beschrijven we de scope van de toets, geven we aan welke ontwerpprincipes getoetst zijn en welke verantwoordelijke partijen betrokken zijn geweest bij deze technische review. In de resultaten worden drie niveaus onderscheiden: compleet, gedeeltelijk compleet en niet compleet. We tonen welke beheermaatregelen compleet zijn bevonden, zie 3.1. In paragraaf 3.2-3.4 zijn de verbeterpunten opgenomen voor de beheersmaatregelen die

gedeeltelijk compleet zijn bevonden en geven wij een advies per beheersmaatregel. In deze review zijn geen beheersmaatregelen niet compleet bevonden. Tot slot vatten we onze bevindingen samen in de conclusie en aanbevelingen.

1 Projectomschrijving

Het algoritme dat de gemeente Amsterdam heeft laten reviewen is het Blurring-as-a-Service (BaaS) algoritme. Dit algoritme heeft als doel om panoramabeelden te ontdoen van persoonsgegevens. In de huidige versie van het algoritme gaat het om directe persoonsgegevens (personen) en herleidbare persoonsgegevens (kentekens). De uitkomst van het algoritme zijn panoramabeelden waarbij deze persoonsgegevens door de toegevoegde blurring niet meer zichtbaar zijn zodat deze hergebruikt kunnen worden door ambtenaren binnen de gemeente. Zij kunnen de beelden gebruiken om bijvoorbeeld onderzoek naar de buitenruimte te doen, zonder deze te betreden.

Het BaaS-algoritme maakt gebruik van een bestaande module voor objectdetectie (YoloV5). Met het detectiegedeelte van deze aangepaste module worden personen en kentekens met bounding boxes geïdentificeerd. Voor de blurring zelf is met een algoritme ruis toegevoegd om de pixels in de gedetecteerde bounding boxes willekeurig te veranderen zodat personen en kentekens niet meer te herkennen zijn of de blur op te heffen. De originele beelden worden niet langer door CVT bewaard dan nodig is voor het blurren.

CVT heeft aan de hand van gesprekken met de interne klant en het CivicAILab groepen geïdentificeerd waar mogelijk discriminatierisico's kunnen spelen. Aan de hand van deze groepen is met behulp van statistieken gezocht naar een gebalanceerde set van foto's om het algoritme te valideren. Voor de training van de objectdetectie zijn afbeeldingen handmatig geannoteerd om zo het algoritme te laten leren hoe het in deze context (panaromafoto's in Amsterdam), personen en kentekens te detecteren. De gemeente Amsterdam heeft er expliciet voor gekozen om een 'inclusief' algoritme te ontwikkelen zodat elke persoon en elke type kenteken evenveel kans heeft om onherkenbaar gemaakt te worden. Om dit bereiken is een iteratief proces gevolgd om de potentieel discriminatoire factoren meetbaar te maken en de resultaten van het algoritme continu te toetsen aan een gebalanceerde validatieset.

2 Scope van deze technische review

Dit hoofdstuk geeft een overzicht van de ontwerpprincipes die zijn getoetst tijdens de technische review per ontwerpprincipe. De ontwerpprincipes en de korte toelichting ervan zijn in de tabel hieronder weergegeven:

Ontwerpprincipes	Aantal maatregelen
Bewust gebruik	15
Gebaseerd op kennis	7
Gecontroleerde toepassing	23
Lerend en evoluerend	5

De documenten gebruikt bij deze toets zijn hieronder in een tabel weergegeven. Daarnaast is een korte beschrijving van de documentatie-inhoud en de bijbehorende verantwoordelijke partij te vinden. Het is mogelijk dat meer dan één partij verantwoordelijkheid draagt binnen een (ander onderdeel van het)

proces.

Documentnaam	Verantwoordelijke partij	Omschrijving
Innovatie budget technische documentatie	Computer Vision team	Omschrijving data; data- en modelkeuzes; statistische onderbouwing; validatie; resultaten
Algoritme source code	Computer Vision team	Broncode van het algoritme
Fairness - risico's en maatregelen	Data scientist binnen het Computer Vision team	Invulling van de Fairness principes
Fairness Handbook	Beleidsmedewerker	Handboek waarin de principes van Fairness en uitwerking worden beschreven
Fairness analyse	Data scientist binnen het Computer Vision team	Invulling Fairness vragen
BaaS Onboarding checklist	Computer Vision team	Checklist om te gebruiken bij het onboarden van nieuwe klanten
BaaS Onboarding proces	Computer Vision team	Procesbeschrijving van onboarding van nieuwe klanten
BaaS Onboarding procesflow v0.1	Computer Vision team	Procesflow van onboarding van nieuwe klanten
BaaS - Monitoring en Evaluatie	Computer Vision team	Beschrijving van de monitoring en evaluatie aanpak BaaS
ORBS BaaS - Service Model Canvas	NA	Beschrijving van het service aanbod BaaS
Particpatieplan project Beeldherkenningsvoorziening v.06 (zonder financiën)	Directeur Digitalisering en Innovatie	Plan voor participatietraject voor ophalen burgerinput over beeldherkenningsvoorziening.

3 Resultaten van de technische review

In dit hoofdstuk worden de aannames en resultaten van de algoritmetoets besproken. Ten eerste volgt een overzicht van aannames en uitgangspunten die naar voren zijn gekomen tijdens het proces van de algoritmetoets. Daarna worden de resultaten geaggregeerd weergegeven per ontwerpprincipe. Vervolgens worden de resultaten meer in detail toegelicht.

De aannames die zijn gemaakt tijdens de ontwikkeling van het BaaS-algoritme zijn in onderstaande opsomming weergegeven:

- De business heeft baat bij geanonimiseerde panoramabeelden en kan en wil deze gaan gebruiken.
- De beelden in de trainingsset zijn representatief en blijven dat ook in de nabije toekomst. De panoramabeelden die gebruikt zijn om het BaaS-algoritme te trainen komen uit 2016 en de jaren daarna. De inhoud en de spreiding van de beelden zijn gebaseerd op schattingen welke beelden bruikbaar zouden zijn. Hierdoor zijn er minder foto's geselecteerd uit de Covid-19 periode, omdat destijds duidelijk minder personen en verkeer op de foto's terug te vinden is.
- Een compleet gebalanceerde trainingsset is niet noodzakelijk als de klassen wel in de validatieset worden herkend. Voor een eerlijke werking van het algoritme voor alle diversiteitscategorieën, is het voldoende dat elke categorie dezelfde kans heeft om onherkenbaar gemaakt te worden.
- De diversiteitscategorieën voor persoonsdetectie zijn te verdelen over leeftijdsgroepen, groepen met verschillende huidskleuren en geslacht.
- Personen die niet identificeerbaar zijn door het menselijk oog zijn voldoende onherkenbaar gemaakt. Dit speelt in gevallen waarbij personen die extreem ver weg zijn van de camera gemist worden in herkenning. Er wordt in de validatie meer gewicht toegekend aan de prestaties in de herkenning van personen die ook daadwerkelijk herkenbaar zijn voor het menselijk oog.
- Buitenlandse kentekens zijn ondergerepresenteerd bij gebrek aan foto's waarop dit type kentekens te identificeren is. De verwachting van de geïnterviewden is dat dit type kentekens tevens ook zelden op nieuwe panaromabeelden te vinden zijn.

De onderstaande lijst beschrijft de uitganspunten die bij de ontwikkeling van het BaaS-algoritme worden gehanteerd:

- Door het werken met panoramafoto's verandert de aspectratio van de beelden. Het algoritme is specifiek getraind op deze 'verstoorde' verhoudingen en is per definitie niet direct geschikt voor niet-panoramafoto's.
- Op basis van statistieken en geschoolde gok, zijn beelden over verschillende locaties verdeeld over de tijd en het gebied voor annotatie verzameld. Deze selectie is iteratief aangepast om voldoende persoonsgegevens, zowel in aantallen als diversiteit, te hebben in training- en validatieset.
- De scanauto's rijden niet gedurende evenementen zoals Koningsdag, waardoor beelden van burgers in uitbundige outfits niet tot de geteste set behoren. Omdat de auto's dan ook niet rijden is het ook minder noodzakelijk deze op te nemen in de trainingsset; er zijn immers minder beelden van burgers in uitbundige outfits.
- De definitie van persoonlijke informatie moet in balans zijn met het verliezen van informatie over de context op de foto's. Oftewel, de bounding box moet zo dicht mogelijk tegen de persoonlijke informatie aan zitten, en tegelijkertijd wel goed genoeg deze informatie maskeren. Deze afweging leidt ertoe dat de bounding box groter is gemaakt met 10% zodat met meer zekerheid persoonlijke informatie wordt afgedekt zonder aanvullend verlies van de context. Bij de ontwikkeling van dit algoritme is persoonlijke informatie beperkt tot wat zichtbaar is van de persoon zelf en kentekens. Context zoals kleding en attributen zijn niet de focus voor de detectie.

De technische review laat zien dat CVT de technische kant van het algoritme nog steeds adequaat heeft ingericht. Er is duidelijk en aantoonbaar nagedacht over mogelijke bias, welk resultaat verwacht wordt en hoe dat resultaat bereikt moet worden. Ook zien we dat het team sinds de vorige review grote stappen heeft gemaakt in het opstellen van voldoende, toegankelijke en meer volledige documentatie. Het monitorings- en evaluatiedocument en het onboardingsdocument zijn een goede aanvulling op de -tevens verbeterde- technische documentatie. Het is nu voor de klant duidelijk wat er van hen verwacht wordt bij onboarding en gedurende de implementatie en tijdens gebruik van BaaS.

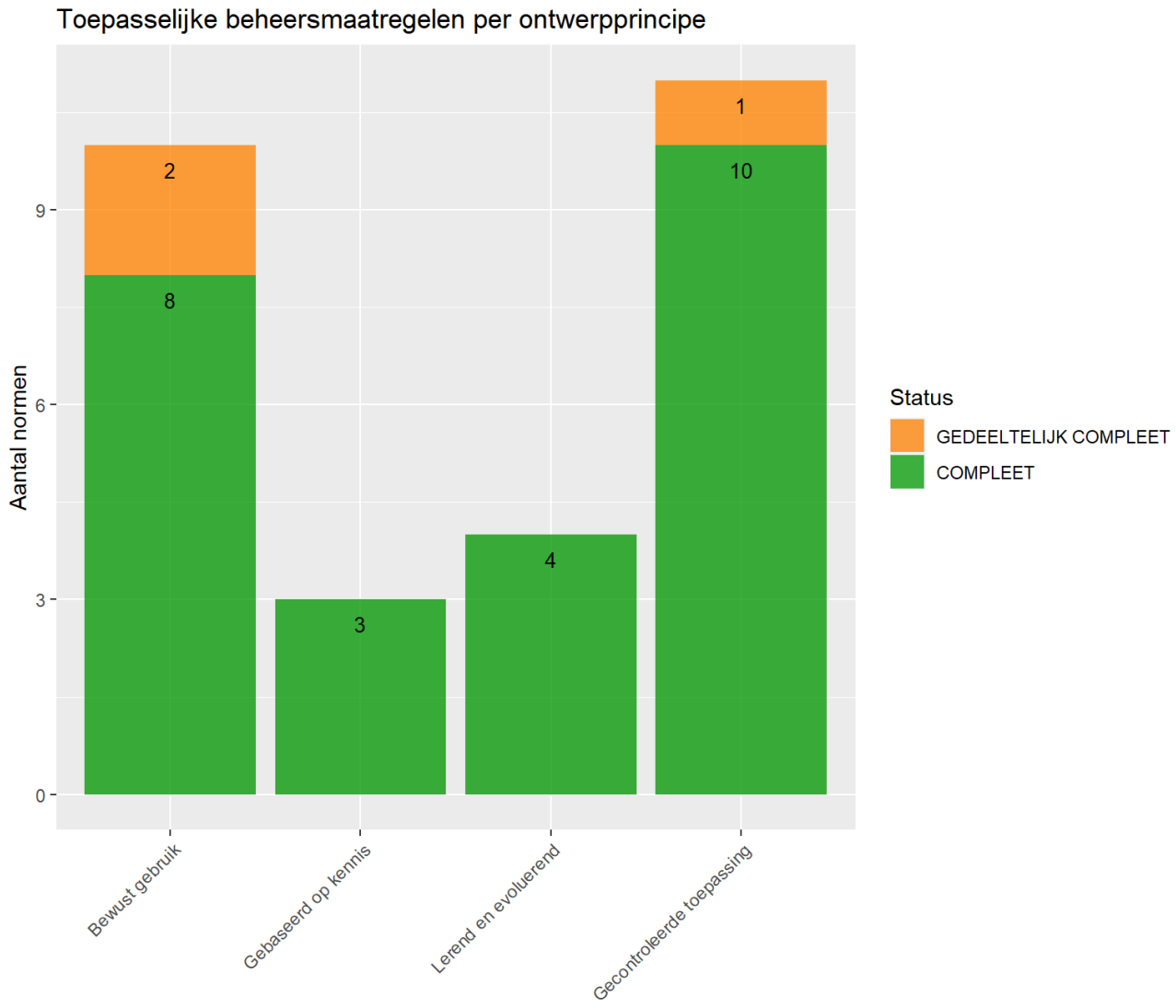
Wel zien we nog verbetermogelijkheden in de duidelijkheid van de documentatie en het beleggen van bepaalde verantwoordelijkheden in het gebruik van BaaS met de klant. Hiervoor hebben we specifieke adviezen opgenomen bij de beheersmaatregelen die gedeeltelijk compleet bevonden zijn (3.2.-3.5.). In deze review zijn alle beheersmaatregelen compleet of gedeeltelijk compleet bevonden. Er zijn geen beheersmaatregelen die in volledigheid niet compleet zijn.

In de documentatie van het particpatietraject worden de ethische vragen genoemd waaraan gedacht kan worden. Het is uit deze documentatie niet duidelijk of de vragen rondom BaaS ook echt onderwerp van gesprek zullen zijn of slechts optioneel. Wij geven ter overweging mee of deze aanpak voldoende is om te bepalen of ook vanuit de burger de inzet van BaaS eerlijk en inclusief genoeg is.

Naar aanleiding van de eerste review zijn door CVT aannames en overwegingen opgenomen in de technishe documentatie. Deze zijn echter ambigu geformuleerd. We hebben de maatregelen technische documentatie als compleet aangemerkt, maar nemen in het technische document opmerkingen op ter aanscherping van de formulering.

De intentie om het juiste goed te doen is duidelijk aanwezig in het team. We zijn er dan ook van overtuigd dat met het opvolgen van de aanbevelingen bij de beheersmaatregelen en aanscherpingen in de bewoording in de technische documentatie CVT voldoende de technische risico’s afdekt voor de inzet van BaaS.

Grafiek 1 hieronder toont de algemene resultaten uit de technische review. In deze grafiek wordt getoond hoeveel toepasselijke beheersmaatregelen per ontwerpprincipe compleet (groen), gedeeltelijk compleet (geel) of niet compleet (rood) bevonden zijn tijdens de technische review.



Grafiek 1: Toepasselijke beheersmaatregelen per ontwerpprincipe

Hier is te zien dat de beheersmaatregelen voor ‘Gebaseerd op kennis’ en ‘Lerend en evoluerend’ allen compleet bevonden zijn. Dit betekent dat we uit deze review concluderen dat het BaaS-algoritme een adequate weergave is van de ervaring van de organisatie. Ook op ‘Bewust gebruik’ heeft CVT al veel stappen gezet. Zij begrijpt de impact van het gebruik van dit algoritme in de praktijk en kan vooral nog stappen zetten in de verslaglegging. Een nadere toelichting op deze bevindingen is te vinden in paragraaf 3.2. Om de maatregelen aangaande ‘Gecontroleerde toepassing’ te completeren is het nodig om ook de fairness richtlijnen te volgen bij een aanvulling of vervanging van de trainingset, zie paragraaf 3.4.

3.1 Complete beheersmaatregelen

CVT heeft een iteratief proces gevolgd om een zo ‘inclusief mogelijk’ algoritme te ontwikkelen. Dit betekent dat personen uit elke categorie evenveel kans hebben om onherkenbaar te zijn op de beelden die worden geblurd. Door de focus te leggen op een evenwichtige validatie is het continu te meten geweest of het algoritme daadwerkelijk voor elke categorie even goed persoonsgegevens onzichtbaar maakt. Deze aanpak leidt ertoe dat de trainingsset telkens is aangevuld met ondergerepresenteerde voorbeelden. Het team is hierdoor in staat geweest een voldoende kwalitatieve set neer te zetten binnen de technische prestatie-eisen. Daarnaast heeft het team na de eerste review de documentatie zo

goed mogelijk aan gevuld om aan te tonen in hoeverre BaaS aan de beheersmaatregelen voldoen. Met de aanpassingen zijn nu meer beheersmaatregelen voldoende compleet bevonden ten opzichte van het eerste reviewmoment.

Het iteratief werken is ook terug te zien in de wijze waarop de data scientisten hebben samengewerkt aan dit algoritme. Veel van de keuzes zijn goed terug te vinden in de tracking van technische wijzigingen in versiebeheertool Github. Door deze tool goed te gebruiken zijn wijzigingen ook altijd door een tweede paar ogen getoetst. Daarnaast wordt elk codecomponent voorzien van een unittest om potentiële fouten tijdens de ontwikkeling al te signaleren. Op het eerste review moment is gebleken dat er parameters direct in de code waren opgenomen en nog niet seperaat gedefinieerd waren. Dit is ten goede aangepast in de laatste versie van BaaS, die beschikbaar is op Github.

Tot slot, is uit de interviews gebleken dat het effect van aanpassingen in de aannames en uitgangspunten, prestatie-eisen en de samenstelling van de data op de kwaliteit van het BaaS-algoritme bij alle geïnterviewden scherp op het netvlies staat.

Ontwerpprincipes	Beschrijving beheersmaatregel	Status
Bewust gebruik	De samenhang met, volgordelijkheid van en belang in het proces waarvan het algoritme onderdeel uitmaakt, is vastgesteld en beschreven.	COMPLEET
Bewust gebruik	De organisatie heeft expliciete fout- en tolerantieniveaus gedefiniëerd (bijv. 60% of 90% nauwkeurigheid).	COMPLEET
Bewust gebruik	Medewerkers van CVT begrijpen de risico’s, kosten en impact van voorspellingsfouten, ook al vallen de fouten binnen de acceptabele fouttoleranties.	COMPLEET
Bewust gebruik	CVT heeft mogelijke ethische bezwaren tegen het gebruik van het algoritme afgewogen bij haar belanghebbenden en betrokkene(n) (= data subjects).	COMPLEET
Bewust gebruik	Het algoritmecreatieproces volgt een structurele aanpak, inclusief volgende onderdelen: - Het gebruik van representatieve training, test en productie data. - Het vaststellen van prestatiedoelen. - Het hebben van een minimale verslaglegging. - Het bepalen van minimaal benodigde gegevens, infrastructuur en kennis.	COMPLEET
Bewust gebruik	Wiskundige model en data aannames zijn gebaseerd op relevante proceskenmerken en toepasselijke business rules.	COMPLEET
Bewust gebruik	Er is een rechtvaardiging voor business rules en proceskenmerken die niet zijn opgenomen in het algoritme.	COMPLEET
Bewust gebruik	De kwaliteit van gebruikte software en packages is getoetst in de praktijk en afgewogen tegen alternatieven in de context van de toepassing.	COMPLEET

Ontwerpprincipes	Beschrijving beheersmaatregel	Status
Gebaseerd op kennis	Het aggregatieniveau van gegevens is adequaat en consistent binnen alle datasets (training data, test data en productie data).	COMPLEET
Gebaseerd op kennis	Het voorspellingsvenster voor gegevenstypen is consistent (test data, training data, productie data) en in overeenstemming met de probleemdefinitie.	COMPLEET
Gebaseerd op kennis	Wiskundige modelaannames en gebruikte data komen overeen met de gegevensstructuur, probleem- en uitkomstdefinitie en zijn in overeenstemming met de toleranties zoals overeengekomen.	COMPLEET
Lerend en evoluerend	Organisatie volgt een kwantitatief proces, waarbij tastbare prestatietesten worden gehanteerd om ervoor te zorgen dat er geleerd wordt van fouten in het algoritme (bijv. een zelflerend mechanisme of door deskundige wijzigingen van het algoritme) en het algoritme continue verbeterd wordt.	COMPLEET
Lerend en evoluerend	De organisatie moet een gestructureerde aanpak hebben om de modellen te valideren en de bijbehorende methoden en resultaten te documenteren in lijn met de doelstellingen. De validatiebenadering moet een validatieschema, een gestandaardiseerd validatierapport en voorgestelde verbeteringen van de datapreparatie én het algoritme bevatten.	COMPLEET
Lerend en evoluerend	Aannames, business rules en prestaties worden periodiek en bij wijzigingen geëvalueerd door het bedrijf en teruggekoppeld aan de algoritmeontwikkelaars. Deze feedback moet ook betrekking hebben op algoritme-resultaten, verbeterpunten en kosten & impact van algoritmisch geïnformeerde beslissingen.	COMPLEET
Lerend en evoluerend	De organisatie volgt een proces om ervoor te zorgen dat er een feedbacklus is om te begrijpen en beheersen hoe interventies uit eerdere analyses het toekomstige gebruik van het algoritme beïnvloeden.	COMPLEET
Gecontroleerde toepassing	Business rules, fouttoleranties (proces en uitkomst) en rekenregels van profielen zijn gedefinieerd en worden alleen beschikbaar gesteld op een need-to-know-basis.	COMPLEET
Gecontroleerde toepassing	De uitkomsten van het algoritme kunnen niet afzonderlijk en / of handmatig worden gewijzigd.	COMPLEET

Ontwerpprincipes	Beschrijving beheersmaatregel	Status
Gecontroleerde toepassing	De verdeling in gevallen tussen training, test en validatieset komt zoveel mogelijk overeen met de echte wereld (voorspellingsset).	COMPLEET
Gecontroleerde toepassing	De organisatie is zich bewust van de biasrisico's en heeft maatregelen genomen om discriminatie door het model te meten, te voorkomen of te corrigeren.	COMPLEET
Gecontroleerde toepassing	De organisatie is zich bewust van de data kwaliteit en de invloed op gemaakte aannames en resultaten. De organisatie heeft maatregelen genomen om de impact op de resultaten te mitigeren.	COMPLEET
Gecontroleerde toepassing	Gevallen in de validatieset buiten de steekproef worden duidelijk gescheiden van de gevallen in de trainingsset volgens de definitie en de stappen voor data voorbereiding.	COMPLEET
Gecontroleerde toepassing	Validatie- en prestatietesten worden voldoende vaak uitgevoerd op algoritmes en zijn samen met monitorings- en signaleringsprocedures ingebed in bedrijfsprocessen.	COMPLEET
Gecontroleerde toepassing	Prestatie-eisen voor het algoritme zijn gedefinieerd en sluiten aan bij de procesdoelstellingen.	COMPLEET
Gecontroleerde toepassing	Algoritme wordt begrijpelijk gecommenteerd, als herinnering aan het doel van de analyse.	COMPLEET
Gecontroleerde toepassing	Algoritme is gestructureerd om onnodige complexiteit te voorkomen en transparantie te creëren.	COMPLEET

In de volgende paragrafen is per ontwerpprincipe een overzicht te vinden van de verbeterpunten en ons advies aangaande het voldoen aan deze beheersmaatregelen. De resultatensectie wordt afgesloten met een overzicht van de compleet bevonden beheersmaatregelen.

3.2 Ontwerpprincipe 1: Bewust gebruik

De ontvangen technische documentatie bevat de gemaakte aannames en uitgangspunten, keuzes en tussentijdse beslissingen. Aanvullend op deze rapportage doen we een suggestie voor de betreffende punten om de intentie van CVT niet-ambigu te verwoorden.

Voor het compleet bevinden van deze maatregelen zijn scherpere afspraken tussen de klant en CVT over de exacte verantwoordelijkheid van beide partijen noodzakelijk. Wij zien nog witte vlekken in het pakken van verantwoordelijkheid als het gaat om onderprestatie van het algoritme en de mogelijk daaropvolgende wijze van beeldenselectie en verbeteren van modelprestaties.

In onderstaande tabel is een overzicht te vinden van de gedeeltelijke complete beheersmaatregelen van het ontwerpprincipe ‘Bewust gebruik’. Daarnaast staan onze bevindingen en aanbevelingen voor verbetering.

ID Maatregel	Ontwerpprincipes	Beschrijving beheersmaatregel	Status	Verificaties, advies en afwijkingen
B1.1.3	Bewust gebruik	Het algoritme levert een continue meetbare bijdrage aan de procesdoelstellingen en voldoet aan de opgestelde kwantitatieve business doelstellingen.	GEDEELTELIJK COMPLEET	De proces- en businessdoelstellingen zijn nog niet bepaald. Deze doelstellingen zijn afhankelijk van de uiteindelijke toepassing van BaaS door de klant. In de interviews is naar voren gekomen dat het algoritme, op basis van een keuzemenu, zoveel mogelijk hierop aansluit. In het onboardingsdocument worden deze toekomstige afspraken over business doelstellingen en prestatie-eisen per toepassing vastgelegd. Per toepassing wordt een monitor en evaluatieproces ingericht. In het voorgestelde onboardingsdocument voorzien wij het risico dat bij herevaluatie de klant afdwingt om lagere prestatie-eisen te accepteren, die niet vooraf zijn overeengekomen bij het doelgebruik. Wij adviseren om in dit soort situaties ‘nee’ te verkopen wanneer niet aan de onboardingseisen voldaan kan worden. Bij het accepteren van de onderpresentatie ontstaat namelijk het risico op o.a. imagoschade voor CVT en BaaS. Als beheersmaatregel kan het team van tevoren een risico-analyse doen op de impact van de mogelijke

ID Maatregel	Ontwerpprincipes	Beschrijving beheersmaatregel	Status	Verificaties, advies en afwijkingen
				oorzaken van onderprestatie. Het resultaat dient dan om de klant te informeren over de uiteindelijke beslissing om wel of niet verder te gaan met de beoogde toepassing van BaaS binnen deze context.

ID Maatregel	Ontwerpprincipes	Beschrijving beheersmaatregel	Status	Verificaties, advies en afwijkingen
T1.3	Bewust gebruik	Het algoritme moet betere resultaten opleveren dan een willekeurige selectie of steekproef op basis van een paar relevante attributen.	GEDEELTELIJK COMPLEET	Er is enkel vanuit haalbaarheidsoptiek gekeken naar de package om te gebruiken voor objectdetectie. Uit interviews blijkt dat niet alleen vanuit de privacy officers, maar ook vanuit technische team het streven is om de nauwkeurigheid te vergroten, idealiter richting de 100%. In de technische documentatie is opgenomen dat het in dit stadium van het ontwikkelproces nog niet loont om alternatieven te testen. De eigenaar van de beelden moet tijdens de onboarding nog definitief wat de foutmarge mag zijn (gelijkwaardig verdeeld over de categorieën). Deze marge zal afhankelijk zijn van de context waarin BaaS gebruikt gaat worden. Het verbeteren van de trainingset specifiek voor de casus van de klant heeft alleen de voorkeur wanneer aan de inclusiviteitsdoelstellingen vanuit CVT voldaan kan worden. Namelijk door aantoonbaar op de aangeleverde testdata van dezelfde casus. Indien het nodig is om de prestaties te verbeteren raden we aan om, naast YoloV8, ook te kijken naar

ID Maatregel	Ontwerpprincipes	Beschrijving beheersmaatregel	Status	Verificaties, advies en afwijkingen
				alternatieven of de parametersettings binnen YoloV5 aan te scherpen. Daarnaast spreekt het voor zich - zoals opgenomen in de MVP - dat door de mogelijkheid om op caseniveau feedback te geven op de gemiste persoonsgegevens, de trainingsset kan worden aangevuld om ook edge cases steeds beter onherkenbaar te kunnen maken.

3.3 Ontwerpprincipe 2: Gebaseerd op kennis

Voor het ontwerpprincipe ‘Gebaseerd op kennis’ zijn alle beheersmaatregelen compleet bevonden.

Het algoritme voldoet compleet aan de toepasselijke beheersmaatregelen onder dit ontwerpprincipe.

3.4 Ontwerpprincipe 3: Lerend en evoluerend

Voor het ontwerpprincipe ‘Lerend en evoluerend’ zijn alle beheersmaatregelen compleet bevonden.

Het algoritme voldoet compleet aan de toepasselijke beheersmaatregelen onder dit ontwerpprincipe.

3.5 Ontwerpprincipe 4: Gecontroleerde toepassing

Het onderwerp ‘Gecontroleerde toepassing’ bevat één beheersmaatregelen die gedeeltelijk compleet bevonden is. Deze maatregel was in de eerste review compleet bevonden. Wij zien echter een risico bij het opnieuw trainen op data vanuit de klant als het gaat om het intensieve proces om een gebalanceerde testset te construeren en de benodigde foto’s te labelen op de potentiële discriminatoire factoren. In onderstaande tabel wordt deze bevinding en aanbeveling verder uitgewerkt.

ID Maatregel	Ontwerpprincipes	Beschrijving beheersmaatregel	Status	Verificaties, advies en afwijkingen
B6.4	Gecontroleerde toepassing	De organisatie heeft mogelijke vooroordelen onderzocht die veroorzaakt kunnen zijn door menselijke subjectiviteit of het gebruik van onjuiste gegevens.	GEDEELTELIJK COMPLEET	In de initiële training van BaaS-algoritme is door CVT in grote zorgvuldigheid een inclusieve trainingset opgesteld. Echter, wordt aan toekomstige klanten gevraagd om bij een andere type foto of onderpresentatie op dit nieuwe formaat foto's aan te leveren. Uit de documentatie wordt niet duidelijk hoe de selectie van beelden voor de testset plaatsvindt voor het annoteren door CVT. Ook is niet bekend hoe gewaarborgd wordt dat de hertraining van het algoritme binnen de inclusiviteitsdoelstellingen blijft en hoe dit voldoende goed te meten is. De huidige leermethode van het algoritme vormt de basis van de aanpak van CVT en de reden om zelf een inclusief algoritme te realiseren. Enkel 100 (test)beelden annoteren volgens de annotatierichtlijnen is niet voldoende om deze claim te waarborgen.

4 Conclusie & aanbevelingen

Dit rapport is het resultaat van de vraag om een re-review uit te voeren op de eerder besproken aanbevelingen.

Het Computer Vision Team (CVT) heeft een iteratief proces gevolgd om een zo 'inclusief mogelijk' algoritme te ontwikkelen. Dit betekent dat personen uit elke categorie evenveel kans hebben om onherkenbaar te zijn op de beelden die worden geblurd. Zij is in de huidige versie van dit algoritme

geslaagd de technische prestatie-eisen voor elke categorie te behalen met uitzondering van ‘buitenlandse kentekens’. De kans dat een buitenlands kenteken voorkomt is aanzienlijk klein en het risico op herkenning minder groot, daardoor wordt de impact van deze uitzondering op een minimum geschat.

Met BaaS is CVT geslaagd een middel te ontwikkelen om op inclusieve en verantwoorde wijze panoramafoto’s te anonimiseren.

De belangrijkste aanbeveling is om verantwoordelijkheden van de klant (afnemer intern of extern) en CVT, duidelijk vast te leggen in de documentatie. In de nieuwe en aangevulde documentatie worden sommige verantwoordelijkheden wel besproken, maar blijft de precieze verantwoordelijkheid ambigu. Wat ons betreft is het voor een geslaagde implementatie van dit algoritme essentieel dat deze verantwoordelijkheden duidelijk belegd zijn en blijven, zodat de verantwoordelijkheid voor de zorgvuldigheid en inclusiviteit van het algoritme ook in de toekomst behouden kan worden.

Tot slot bevelen we aan om de ambiguïteit in de technische documentatie tot een minimum te beperken en te sturen op voldoende aandacht voor de aanpak en doel van BaaS tijdens het participatietraject.