Addendum DPIA Verzamelen brondocumenten t.b.v. afhandelen Corona WOB-verzoeken versie 2.0

Behorende bij de DPIA Verzamelen brondocumenten t.b.v. afhandelen Corona WOB-verzoeken versie 21-03-2022

Den Haag, 30 augustus 2022 / versie 1.0 / Status: Definitief

Vaststelling verwerkingsverantwoordelijke:
Functie: 5.1.2e Programmadirectie Nafase Covid-19

Kennisgenomen van FG-advies 12 april 2022

Acties naar aanleiding van het advies van de FG

- Overleg 5.1.2e – CISO & PO
- Uitwerking 5.1.2e en verdere uitvraag bij uitvoeringsteam
- Uitwerking CISO/PO office
- Document 'Nadere uitwerking en aanscherping n.a.v. FG Advies 12 april 2022 – DPIA Addendum Machine Learning Project Behorend bij de DPIA verzamelen brondocumenten t.b.v. afhandelen Corona Wob-verzoeken'' opgesteld.
- Addendum aangevuld

Datum akkoord: d.d. 06-09-2022

5.1.2e



Inleiding

Aanleiding

Dit addendum is opgesteld door het programma 'NC-19' binnen het Ministerie van Volksgezondheid, Welzijn en Sport (hierna: VWS) en is een addendum op de DPIA 'Het verzamelen van brondocumenten voor het afhandelen van Corona gerelateerde Wob-verzoeken' (hierna: de DPIA).

In voornoemde DPIA wordt het proces geïntroduceerd om door middel van "Zoek & Vind" geautomatiseerd bestanden en brondocumenten te (door-)zoeken, met behulp van daarvoor ingehuurde (externe) informatiespecialisten.
Het ministerie van Volksgezondheid, Welzijn en Sport (hierna: min. VWS) ontvangt jaarlijks tientallen verzoeken in het kader van de Wet Openbaarheid van Bestuur (Wob) (hierna: Wob-verzoeken). In 2020 is dit aantal enorm toegenomen door Corona-gerelateerde WOB-verzoeken, door onder andere pers en media. Door het grote aantal verzoeken en beperkte mankracht i.c.m. de wettelijke afhandelingstermijnen voor de Wob, dreigt min. VWS hierdoor in de knel te komen.
Om tijdig te kunnen ageren is hierop vanuit de Directie Organisatie Bedrijfsvoering en Personeel (OBP) het proces van de DPIA verzamelen van brondocumenten voor het afhandelen van Corona gerelateerde Wob-verzoeken ingevoerd.

Dit addendum ziet op het proces Machine learning project.
Machine learning (ML) is een vorm van Artificial intelligentie (AI) die is gericht op het bouwen van systemen die van de verwerkte data kunnen leren of data gebruiken om beter te presteren. Met Machine learning is de technologie in staat zichzelf te ontwikkelen door data-analyse en vervolgens zelf beslissingen te nemen die niet zijn voorgeprogrammeerd. Machine learning-technologie maakt daarvoor gebruik van zelflerende algoritmes. Met het Machine learning project worden de documenten die voor de Wob juristen het meest relevant zijn voorgeselecteerd.
Als bij de ontwikkeling en/of toepassing van een algoritme persoonsgegevens worden verwerkt, is de Algemene Verordening Gegevensbescherming (AVG) van toepassing. Het toepassen van Machine Learning voorafgaande aan de verwerkingen beschreven in de oorspronkelijke DPIA zijn onderwerp van dit addendum.

Reikwijdte Addendum

De eigenaar van het proces dat deze Addendum oplevert is de 5.1.2e OBP. Het afhandelen van WOB-verzoeken in het algemeen valt onder PDO, waar de 5.1.2e PDO de eigenaar van is. De datahouders zijn de directeuren van de directies, verantwoordelijken voor de programma's en de projecten waar de brondocumenten vandaan komen. De overige betrokkenen en verantwoordelijken bij dit proces zijn de CISO Kern, die controle op het waarborgen van privacy en informatiebeveiliging uitvoert. Zaken als het rapporteren van privacy- en beveiligingsincidenten, diensten van derden en toezien op vervulling en uitvoering van rollen en verantwoordelijkheden uitgevoerd onder het toezien van de projectleider van het project dat brondocumenten verzamelt voor Corona gerelateerde Wob-verzoeken. De projectleider werkt onder de verantwoordelijkheid van het afdelingshoofd van het CIO-office van de Directie OBP.

Een extra DPIA of wijziging van de DPIA is niet nodig is maar een aanvulling wel. Een duidelijke procesbeschrijving waarbij de privacy en informatie beveiligingsaspecten goed zijn geborgd, is raadzaam. De AVG legt de lat behoorlijk hoog. Het is dan ook van belang dat een goede analyse wordt gemaakt van alle vormen van verwerking. Dat biedt een goede basis voor het nemen van gemotiveerde besluiten.

De aanvullingen op de DPIA zijn: H3 (procesbeschrijving), H8 (methodiek en methoden) en H13 (proportionaliteit).

Binnen de reikwijdte van dit addendum op de DPIA valt:

- Aanvullingen op basis van de toevoeging van het gebruik van machine learning tijdens het selecteren van documenten die relevant zijn voor de beantwoording van Wob-verzoeken. Dit omvat met name:
    - Uitbreiding van het proces (Paragraaf 3) doelverbinding (Paragraaf 4) locatie nieuwe server (Paragraaf 7)
    - Uitleg omtrent de techniek en methode van de gegevensverwerking. (Paragraaf 8)
    - Afweging of de extra methode impact heeft op de noodzaak en proportionaliteit van de gegevensverwerking (Paragraaf 13) Zorgt de extra methode voor meer risico's? En zo ja, welke maatregelen worden hiervoor getroffen? (Onderdeel C en D)

Buiten reikwijdte van dit addendum op de DPIA valt:

- Dit addendum betreft niet het beoordelen van Wob-verzoeken, noch de uiteindelijke afhandeling ervan door de Wob-jurist.

A Beschrijving kenmerken gegevensverwerking

1 Voorstel

1.1 Project

De verwerkingsverantwoordelijke heeft in de Wet Openbaarheid van Bestuur (Wob) een wettelijke plicht om aan Wob-verzoeken te voldoen. Op grond van artikelen 3 en 8 uit deze wet is de Minister verplicht om indien verzocht informatie over bestuurlijke aangelegenheden openbaar te maken, behoudens de in die wet genoemde uitzonderingen en beperkingen.

Het aanleveren van brondocumenten om te kunnen voldoen aan Wob-verzoeken dient volledig, correct en tijdig te gebeuren. Dit is tevens weer in het belang van het functioneren van de rechtstaat, door het leveren van transparantie aan de maatschappij over bestuurlijke besluiten.
Het ministerie van VWS ontvangt vanwege zijn centrale rol in de Corona pandemie, welke op zijn beurt een zeer grote impact heeft op de samenleving, een groot aantal Corona gerelateerde Wob-verzoeken. De verwerkingsverantwoordelijke heeft hierdoor een belang om volledig, correct en tijdig invulling te geven aan het beantwoorden van de Corona gerelateerde Wob-verzoeken. Het project speelt een essentiële rol om dit te bewerkstelligen.

Het doel van het project is om door middel van machine learning, een rangschikking van documenten te maken. Vervolgens zullen deze documenten op basis van deze rangschikking gepresenteerd worden aan de juristen, waarna zij de documenten op dezelfde manier zullen beoordelen als de huidige situatie. Het doel van dit rangschikken van documenten, is om de meest relevante documenten als eerste bij de wob-juristen te krijgen, om zo de termijn van beantwoording te versnellen. Het gaat hierbij dus enkel om de volgordelijkheid van documenten om te beoordelen te veranderen, dit proces zal geen documenten uitsluiten.

1.2 Procesbeschrijving Machine Learning Project

Onderstaand een beschrijving op hoofdlijnen van het proces.

1. Allereerst zullen we een model trainen op de dataset.
2. Na het trainen zal het model grondig worden geëvalueerd, om te zorgen dat het model doet waarvoor het is gemaakt.
3. Het model zal nieuwe documenten gaan classificeren op basis van een score die de kans weergeeft dat een document onder het verzoek valt.
4. Op basis van de score van de classificatie zal een rangschikking worden gemaakt. Documenten met de grootste kans op 'valt onder verzoek' (dat wil zeggen, met de hoogste score van het model), zullen als eerste worden aangeboden aan de wob-juristen.
5. Wob-Juristen behandelen de ontvangen documenten exact hetzelfde als nu, en zullen beoordelen of deze daadwerkelijk onder verzoek vallen of niet.
6. Wijze van evaluatie

Het doel van het model is het voorspellen of een document wel of niet onder het Wob-verzoek valt. Het model wordt getraind op een dataset van al gelabelde documenten die door juristen zijn beoordeeld. Dit zijn allemaal Corona-gerelateerde documenten en kunnen persoonsgegevens bevatten. Het model zal geëvalueerd worden op al gelabelde documenten van juristen die niet meegenomen zijn in het trainen van het model. Het gaat hierbij dus enkel om de volgordelijkheid van documenten om te beoordelen te veranderen, dit proces zal geen documenten uitsluiten. Er zijn geen additionele privacy risico's.

2 Persoonsgegevens

Er is geen aanpassing of aanvulling ten opzichte van de oorspronkelijke DPIA.

3 Verwerkingen van persoonsgegevens (ADDENDUM TOEVOEGING)

De verwerkingen van persoonsgegevens onder verwerkingsverantwoordelijkheid van VWS wijken niet af van de verwerkingen zoals opgenomen in de oorspronkelijke DPIA.
Er is wel een techniek toegevoegd in een deelproces van verwerkingen. Zie onderdeel P2 - De informatiespecialisten van OBP analyseren het verzoek en verzamelen de relevante documenten uit de bronnen.

4 Verwerkingsdoeleinden

4.1 Verwerkingsdoeleinden

Persoonsgegevens mogen in beginsel niet worden verwerkt voor een doel dat onverenigbaar is met het doel waarvoor de persoonsgegevens oorspronkelijk zijn verkregen.

De data is niet voor de ontwikkeling van algoritmes verkregen. Het ontwikkelen van algoritmes is ook niet het doel. Het doel blijft – overeenkomstig hetgeen in DPIA is opgenomen - relevante documenten aanleveren voor Wob-verzoeken.

Is er voor de verwerking van persoonsgegevens een grondslag en, als sprake is van bijzondere persoonsgegevens, een doorbrekingsgrond aanwezig? Bij AI-toepassingen zal steeds gekeken moeten worden naar de context waarin van AI gebruik wordt gemaakt en welk doel dat gebruik dient. Dat er bijzondere persoonsgegeven in de data voorkomt, is niet 100% uit te sluiten. Graag verwijzen we naar de DPIA voor de genomen maatregelen.

4.1.1 Maatregelen indien gevoelige gegevens worden gevonden

Er worden documenten gebruikt waarbij al gevoelige persoonsgegevens zijn weggelaten. Vanwege de omvang van de dataset, kun je nooit 100% garanderen dat er nog gevoelige persoonsgegevens zich in de dataset bevinden. De uitkomst van het model is alleen ter ondersteuning van de Wob-jurist. Het model zal dus nooit een definitieve beslissing maken. Een Wob-jurist doet dus nog een check en kan eventueel nog gevonden gevoelige gegevens verder lakken.

5 Betrokken partijen

SSC ICT is verwerker onder de AVG. SSC ICT heeft toegang tot de server en onderhoudt de server. VWS-NC19 is verwerkersverantwoordelijke. Team Informatieverzoeken heeft rechten voor deze bestanden. Alle teamleden zijn externe inhuur maar werken onder verantwoordelijkheid van VWS.

6 Belangen bij de gegevensverwerking

De belangen bij gegevensverwerking komen overeen met de belangen zoals beschreven in de DPIA.

7 Verwerkingslocaties

De verwerkingslocaties komen, wat betreft de verwerking door de VWS-servers (die zich in Nederland bevinden), overeen met de locaties zoals in de DPIA beschreven. Er is wel een nieuwe server (omgeving)

aangevraagd bij SSC-ICT. De server heeft geen internet toegang. De server wordt beheerd door SSC-ICT, en het gaat om een Windows Server. Daarnaast is er een beperkte toegang tot de server. Slechts een deel van het team heeft toegang tot deze server, personen die niet mee zullen werken aan de modellen, hebben geen toegang tot de server.

Wijze waarop de data op de server met de software voor machine learning wordt geplaatst
We hebben de mogelijkheid om de data te verplaatsen vanaf de gezamenlijke schijf naar de nieuw beheerde server. Vervolgens wordt het van beheer-server naar de Windows-server verplaatst. Het gaat daarbij om een directe verbinding. Verplaatsing gebeurt middels intern netwerk waar geen internet voor nodig is.
Dit brengt geen aanvullende privacy risico's met zich mee.

- Het model ondersteunt enkel en vervangt niets.
- Er wordt geen gebruik gemaakt van de gelakte versie, maar van de schone versie. De dataset bestaat uit Corona-gerelateerde documenten die mogelijk onder een WOB-verzoek kunnen vallen.
- De dataset is opgesteld van documenten uit ZyLAB. Hierbij wordt een waarschijnlijkheidspercentage aan ieder document gekoppeld en weer terug gestopt als metadata veld in ZyLAB. De resultaten van het proces, zal een score tussen 0 en 1 zijn voor elk document. Dit getal zal aangeven hoe groot de kans is dat een document 'valt onder verzoek' als beoordeling krijgt (volgens het model). Deze scores worden toegevoegd binnen ZyLab als extra metadata veld.

8 Techniek en methode van gegevensverwerking

Onderstaand wordt, voor zover nog niet uitgewerkt in voorgaande paragrafen en ter aanvulling op de DPIA, beschreven op welke wijze en met gebruikmaking van welke (technische) middelen en methoden de persoonsgegevens worden verwerkt bij gebruik van Machine Learning. We geven uitleg over hoe de selectie op basis van machine learning precies werkt. Tevens wordt benoemd of sprake is van (semi-)geautomatiseerde besluitvorming.

8.1 Is sprake van (semi-)geautomatiseerde besluitvorming?

Nee. De selectie op basis van Machine Learning gaat als volgt. De resultaten die worden geproduceerd door het model (een score tussen 0 en 1), zal worden gebruikt om de documenten te rangschikken. Vervolgens zullen de documenten met de hoogste score als eerste worden gepresenteerd aan de juristen, zodat zij de (waarschijnlijk) relevante documenten als eerste te zien krijgen en beoordelen, om zo het proces te versnellen. Het is dus geen geautomatiseerde besluitvorming, omdat alle documenten uiteindelijk alsnog door de jurist worden behandeld, echter de volgorde waarin wordt wel veranderd.

8.2 Beveiliging

In Deel C is een meer uitgebreide beschrijving opgenomen van de (voor)genomen maatregelen

9 Bewaartermijnen

Voor de persoonsgegevens die worden ingevoerd geldt een zelfde bewaartermijn als de oorspronkelijke DPIA. De documenten zijn exact dezelfde documenten, dus ze worden even lang bewaard. Wel zal wanneer het model is gemaakt, de kopie van de data van de server worden gehaald. Echter is de data dan nog steeds beschikbaar op de oorspronkelijke locatie.

Met inachtneming van de beginselen van noodzakelijkheid en dataminimalisatie zullen wij alleen de data bewaren indien nodig en voor zolang nodig. Er is hier geen wijziging ten opzichte van de oorspronkelijke data.

B Beoordeling rechtmatigheid gegevensverwerkingen

10 Rechtsgrond

De rechtsgrond voor de gegevensverwerking wijkt niet af van de rechtsgronden zoals beschreven in de DPIA. De grondslag is zodoende het uitvoeren van een taak van algemeen belang zoals bedoeld in artikel 6 lid 1 sub e AVG. Onderstaand wordt dit nader uitgewerkt.

10.1 Grondslagen

In beginsel is testen met persoonsgegevens niet toegestaan. De AVG legt de verwerkingsverantwoordelijke de verplichting op om aan te tonen dat de verwerking zoveel mogelijk wordt beperkt én dat er zoveel mogelijk passende waarborgen worden getroffen (privacy by design en privacy by default). Alleen indien er aantoonbaar geen alternatief is, is verwerking toegestaan.

Er wordt getest of de geconfigureerde parameters van het machine learning daadwerkelijk de meest relevante documenten naar voren haalt voor de Wob-verzoeken.
De test is daarmee niet op document niveau en niet op persoonsgegevensniveau.

Het gebruik van de machine learning vindt plaats op de dataset waarop de selectie door de Wob-juristen nu ook zou plaatsvinden. Het lakproces volgt daarna. Machine learning wordt toegevoegd om het huidige proces te versnellen. Op basis van beproefde algoritmen worden documenten op relevantie geordend.

11 Bijzondere persoonsgegevens

Er is geen grondslag voor het verwerken van bijzondere persoonsgegevens. Daarbij zijn bijzondere persoonsgegevens ook niet noodzakelijk om het doel te bereiken. Er worden maatregelen genomen om bijzondere persoonsgegevens zoveel als mogelijk buiten de dataset te houden. Zie oorspronkelijke DPIA.

12 Doelbinding

Er is sprake van verdere verwerking in de zin van artikel 6 lid 4 AVG.

De aanvullende verwerking dient dus getoetst te worden op proportionaliteit en subsidiariteit.
- Is het belang van de verantwoordelijke gerechtvaardigd en voldoende zwaarwegend?

Het belang om te voldoen aan de wettelijk termijn van de Wob voor het afhandelen van Wob-verzoeken is voldoende zwaarwegend.

- Zijn alle eventuele alternatieven voor het doel ontoereikend?
Hierbij kan worden gedacht aan pseudonimisering of het gebruik van 'synthetische data'. Er zijn diverse tools ontwikkeld die fictieve datasets genereren. Kunnen jullie aantonen dat deze hulpmiddelen ontoereikend zijn?

Machine learning is een instrument om het doel effectiever te bereiken, namelijk de relevante brondocumenten voor Wob-verzoeken verzamelen.
Dat doel kan niet bereikt worden met synthetische data, die is namelijk per definitie niet relevant voor Wob-verzoeken.

Machine Learning is beproefde techniek. Het model moet worden getraind specifiek op deze dataset. Het is geen experimentele techniek, maar het moet goed werken op onze dataset.

De belangen van de betrokkenen zijn voldoende gewaarborgd.
De selectie vindt niet op persoonsgegevens plaats. Het machine learning model voorspelt niet eigenstandig.

Er zal in dat geval aan de hand van de criteria in artikel 6 lid 4 AVG moeten worden beoordeeld of sprake is van verenigbare doelen. Een verdere verwerking met het oog op onder meer wetenschappelijke onderzoek/ontwikkelingen, waar technologische ontwikkelingen ook onder kunnen vallen, als een verenigbare rechtmatige verwerking wordt beschouwd. Vanzelfsprekend moet het wel aan bepaalde criteria voldoen. Het gerechtvaardigd belang van de verwerkingsverantwoordelijke moet worden aangetoond waarbij verwantschap met het oorspronkelijk doel een voorwaarde blijft. (overweging 50 AVG)

Wob-verzoeken moeten binnen de wettelijke termijn worden afgehandeld. VWS is al op de vingers getikt voor het niet op tijd voldoen aan de wettelijke Wob-termijnen. Dit is een verdere inspiratie geweest om technieken toe te passen die het proces kunnen versnellen.
Zonder dit project wordt het lastig om aan de termijnen te gaan voldoen. VWS wil binnen de wettelijke termijnen aan Wob-verzoeken kunnen voldoen. Door gebruik te gaan maken van Machine Learning zijn we in staat om de meest relevante documenten te selecteren om te laten beoordelen door de wob-jurist. Dat scheelt veel tijd en uitzoekwerk waarvoor geen mankracht is.

Het doel van de verwerking blijft hetzelfde. Alleen de methode wijzigt. Om de methode te wijzigen wordt het eerst getest. Het oorspronkelijk doel blijft ook hier hetzelfde. Meer specifiek heeft de verwerking in dit geval ook als doel om te beoordelen of de methode goed werkt. Relevante documenten voor Wob-verzoeken worden door het gebruik van de machine learning tool gecategoriseerd als niet-relevant en worden daardoor uitgesloten voor het beantwoorden van het Wob-verzoek.

13 Noodzaak en evenredigheid

13.1 Proportionaliteit

De vraag die moet worden beantwoord is of de inbreuk op de persoonlijke levenssfeer en de bescherming van de persoonsgegevens van de Betrokkenen in evenredige verhouding staan tot de verwerkingsdoeleinden. Voor de beoordeling van de proportionaliteit van de voorgenomen verwerking zijn op basis van de jurisprudentie van het EHRM en de opinies van de WP29 (de voorganger van de EDPB) de belangrijkste criteria:

a) Niet in strijd met een wettelijke bepaling;
b) Een legitiem doel nastreven;
c) Noodzakelijk in een democratische samenleving.

a. Niet in strijd met een wettelijke bepaling
Wij hebben geen wettelijke bepaling geconstateerd die in strijd is met de voorgenomen verwerking.

b. Een legitiem doel nastreven
Het doel wat nagestreefd wordt met dit project wordt haalbaarder en komt voort uit een wettelijke taak van de Rijksoverheid.

Waarbij criterium c, de noodzakelijkheid in een democratische samenleving, getoetst wordt aan de hand van deelcriteria:

- Dringende maatschappelijke noodzaak;
- Proportionaliteit van het middel voor het doel;
- Relevante en toereikende redenen.

c. Noodzakelijk in een democratische samenleving

Dit criterium valt uiteen in drie subcriteria:

- Dringende maatschappelijke noodzaak:

Er is sprake van een uitdrukkelijke politieke wens om de grote hoeveelheid wob-verzoeken die tijdens de coronaperiode zijn binnengekomen, tijdig af te handelen.

- Proportionaliteit van het middel voor het doel:

Dit doel is helder en het middel is passend om dit doel te bereiken.

- Relevante en toereikende redenen:

De verwerking is nodig om de documenten goed te kunnen ordenen.

13.2 Subsidiariteit

De belangrijkste subsidiariteitsvraag is of de verwerkingsdoelen bereikt kunnen worden met een minder vergaande verwerking van persoonsgegevens.

Uitgangspunt is dat de verwerking van persoonsgegevens tot een minimum moet worden beperkt.
Het gegevensgebruik moet zo minimaal mogelijk en toegespitst op het doel van de verwerking.

Er is sprake van dataminimalisatie want er worden niet meer persoonsgegevens verwerkt dan noodzakelijk. De originele data is nodig, net zoals in de huidige procedure. Met de machine learning methode worden er naar verwachting niet meer persoonsgegevens verwerkt dan met de huidige selectiemethode.

Voor de Proof of Concept (POC) worden 15.000 documenten gebruikt. Dit aantal documenten is nodig om de haalbaarheid aan te kunnen tonen.

De doeldocumenten zullen sowieso worden verwerkt en het is dus geen extra verwerking. Echter, dat de doeldocumenten worden verwerkt wil niet zeggen dat er geen sprake kan zijn van een ander soort verwerking t.a.v. deze documenten.

De inschatting van het team OBP informatieverzoeken is dat we met het trainen van een model van 15.000 documenten een accuraat model kunnen bouwen.

We zorgen er ook voor dat de persoonsgegevens die worden verwerkt juist zijn. We zorgen voor een juiste kopie van de originele dataset.
Dit geldt ook voor nieuwe gegenereerde data uit bestaande datasets. Zo dwingt het juistheidbeginsel tot het beperken van ´hidden biases´ in input en gegenereerde output.

We zullen voor het trainen van de dataset een deel van onze huidige dataset kopiëren en hergebruiken. De data die we gebruiken is al beoordeeld door de Wob-juristen juristen (we hebben hun gegeven oordeel nodig om het model te trainen). Hierdoor weten we dat deze data accuraat is en niet meer zal veranderen, want wanneer een oordeel is gegeven, zal dit niet meer worden aangepast.

14 Rechten van Betrokkenen

De wijze waarop navulling wordt gegeven aan de rechten van de Betrokkenen wijkt niet af van wat in de DPIA wordt aangegeven.

C Beschrijving en beoordeling risico's voor de Betrokkenen, (voor)genomen maatregelen en restrisico's

15 Risicoanalyse

15.1 Risicobegrip

In dit onderdeel worden de risico's van de voorgenomen gegevensverwerkingen voor de rechten en vrijheden van Betrokkenen die worden voorzien beschreven en beoordeeld. Hierbij wordt, conform de richtlijnen van WP29 voor gegevenseffectbeoordelingen[1], uitgegaan van een intrinsiek risicobegrip. Daarmee wordt bedoeld dat de kans en de impact van de negatieve gebeurtenis beoordeeld worden zonder dat eventuele reeds genomen maatregelen meegenomen worden in de beoordeling.

In aanvulling op de risico's uitgewerkt in de DPIA, betreft het ten aanzien van dit addendum gebeurtenissen die kunnen leiden tot met name:

Risico	R.1	Relevante documenten voor Wob-verzoeken worden door het gebruik van de machine learning tool gecategoriseerd als niet-relevant en worden daardoor uitgesloten voor het beantwoorden van het Wob-verzoek.			
Proces	P.2	Toepassen van machine learning om de meest relevante documenten voor het Wob-verzoek te selecteren.			
Persoons-gegevens	☒ gewone ☒ gevoelige ☐ bijzondere ☐ strafrechtelijke				
Schade	☐ lichamelijke schade ☐ materiële schade ☒ immateriële schade De aanvrager van het Wob-verzoek kan het ministerie aanklagen voor het niet leveren van alle relevante informatie.				
Oorzaken	De oorzaak van deze dreiging is dat de algoritmen niet goed gecategoriseerd zijn en daardoor niet de juiste documenten selecteren.				
Risiconiveau	Impact	x	Kans	=	Risicoscore	Inschatting risico
	☒ klein(1) ☐ middel (2) ☐ groot (3)		☒ laag (1) ☐ middel (2) ☐ hoog (3)		☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 6 ☐ 9	☒ **Risico zeer laag** ☐ **Risico laag** ☐ **Risico middel** ☐ **Risico middel** ☐ **Risico hoog** ☐ **Risico zeer hoog**

Maatregel	**Omschrijving maatregel**	**Beheerder maatregel**
M.1	Handmatige controle op relevantie van de documenten.	

Restrisico's	**Omschrijving restrisico**
-	Er is geen restrisico.

[1] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

D. Beschrijving voorgenomen maatregelen

In onderdeel D wordt bezien welke maatregelen kunnen worden getroffen om de in onderdeel C erkende risico's te voorkomen of te verminderen. Welke maatregelen in redelijkheid worden getroffen is een belangenafweging van de wetgever of verwerkingsverantwoordelijke.

Nadat voor de belangrijkste risico's beheersmaatregelen zijn geïmplementeerd, blijven er mogelijk een aantal "restrisico's" over. De organisatie moet een besluit nemen hoe hier mee om te gaan.
De volgende maatregelen zijn onderkend per risico (inclusief eventuele restrisico's):

Een risico wat zich eventueel voor zou kunnen doen, is dat juristen na verloop van tijd het model blindelings zouden gaan vertrouwen: Wanneer ze merken dat het model valt onder verzoek juist classificeert, zouden ze blindelings het oordeel van het model kunnen overnemen zonder dit zelf te checken.

1. Maatregelen om dit te voorkomen zouden kunnen bestaan uit het volgende:
- Juristen niet op de hoogte brengen van deze nieuwe rangschikking, op deze manier zouden ze niet beseffen dat ze de data op een specifieke manier krijgen gepresenteerd.
- Een aantal documenten met een grote kans 'valt niet onder verzoek' bewust hoger in de rangschikking plaatsen, zodat de juristen er niet vanuit kunnen gaan dat alles wat ze krijgen een grote mogelijkheid heeft om 'valt onder verzoek' te hebben.

1. Risico's, voorgenomen maatregelen en restrisico's
   - voor de betrokkene met betrekking tot overdracht en teruggave op en van de separate server

   De documenten worden gesleept. Dit gebeurt via een beveiligde verbinding. Wij constateren hierbij geen privacy risico's. Er kunnen hierbij geen documenten verloren raken en/of in verkeerde handen belanden.

   Er is dus geen sprake van een onopzettelijk of ongeoorloofd verlies van toegang tot persoonsgegevens of een onopzettelijke of ongeoorloofde vernietiging van persoonsgegevens (inbreuk op de beschikbaarheid of vertrouwelijkheid van de persoonsgegevens). Er kan ook geen sprake zijn van verlies van beschikbaarheid en/of dat de normale dienstverlening ernstig wordt verstoord

   - bij de evaluatie van het model ten opzichte van de wijze waarop de evaluatie plaatsvindt in relatie tot verwerking van persoonsgegevens

   Wij constateren hier ook geen privacy risico's omdat het machine learning methode niet meer of andere persoonsgegevens verwerkt dan de huidige document selectie procedure. De verwachting is dat de methode efficiënt een voorselectie kan maken. Het model zal worden gemaakt + getraind + geëvalueerd worden met Python (programmeertaal). De input en output komen uit en gaan naar ZyLab.