Gemeente Amsterdam

Template Kunstmatige Intelligentie Impact Assessment

V1. Gemeente Amsterdam

Kunstmatige Intelligentie (hierna KI) :
Directie:
Verwerkings-
verantwoordelijke:
Leverings-
verantwoordelijke:
Afdeling:
Afdelingshoofd:
Werkproces:
Proceseigenaar:
Opgesteld door:
Datum:
Versie:

Bijlagen:

1) DPIA werkproces;
2) Excelbestand met overzichten van:
    - Verwerkte data (met doelbinding, relevantie voor ontwikkeling KI);
    - Opgestelde features;
    - Business rules betreffende werkproces;
    - Verwijderde features;
3) Excelbestand met weging betrokken (grond)rechten en waarden inzet KI;
4) Baselinetoets / DRA
5) Rapport [redacted] analyse;
6) Auditrapport;
7) CPA/ FG advies/ adviezen;
8) Collegebesluit;
9) Gebruikersinstructie
10) Beheerplan

Inhoudsopgave:

1. Inleiding p.
2. Bepalen noodzaak uitvoeren Kunstmatige Intelligentie Impact Assessment. p.
3. KI-toepassing, verantwoordelijkheden en werkproces p.
4. Beschrijf de baten van de KI-toepassing toepassing p
5. Architectuur p.
6. Data en data governance
7. Wettige en ethische toepassing p.
8. [redacted] analyse p.
9. Technische documentatie p.
10. Risicomanagement p.
11. Informatiebeveiliging p
12. Beheer p
13. Communicatie p.
14. Verantwoording p.
15. Afweging en beoordeling p.

1. Inleiding

2. Bepalen noodzaak voor het doen van een kunstmatige intelligentie impact assessment

2.1 ***Wordt de kunstmatige intelligentie (hierna: KI) toegepast in een nieuw (maatschappelijk) domein?***

2.2 ***Wordt gebruik gemaakt van een nieuwe vorm van KI-technologie?***

2.3 ***Is er sprake van een verboden KI-toepassing zoals beschreven in art. 5 van de (concept) regelgeving European Approach for AI van de Europese Commissie?***

2.4 ***Heeft de KI een hoge mate van autonomie?***

2.5 ***Wordt de KI toegepast in een complexe omgeving?***

2.6 ***Wordt de KI toegepast binnen een 'gevoelig' werkproces?***

2.7 ***Heeft de KI eerder schade toegebracht aan fundamentele rechten van natuurlijke personen of groepen?***

2.8 ***Zijn er zorgen over de KI en de potentiële schade en impact?***

2.9 ***Wordt gebruik gemaakt van gevoelige gegevens over personen?***

2.10 ***Wordt een voorspelling gegenereerd op persoonlijke niveau op basis van persoonlijke kenmerken?***

2.11 ***Neemt de KI-beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?***

2.12 ***In hoeverre is de output van de KI van invloed op het besluitvormingsproces binnen de organisatie?***

2.13 ***Is er, mede bezien vanuit de geconstateerde en voorzienbare risico's, sprake van een hoge risico KI-systeem?***

2.14 Indien er sprake is van een (potentiële) hoge risico KI-toepassing, is de Autoriteit Persoonsgegevens geïnformeerd over de intentie om de KI-toepassing te ontwerpen en te ontwikkelen?

2.15 Conclusie (uitvoeren KI-assessment of niet?)

Ja? Ga door naar 3. Nee? Sla de afwegingen hierboven inclusief conclusie op.

3. KI-toepassing, verantwoordelijkheden en betreffende werkproces

3.1 ***Beschrijf de toepassing en het doel van de toepassing.***

3.2 ***Beschrijf het werkproces (BPMn uitwerking en toelichting).***

3.3 ***Welke processtap/ taak wordt ondersteund met de KI?***

3.4 ***Worden bij deze processtap/ taak beslisregels toegepast door medewerkers of systemen?***

3.5 ***Welke verantwoordelijkheden zijn er als het gaat om het verwerken van de gegevens?***
Zie document 'taken en verantwoordelijkheden bij het gebruik van Algoritmen en rekenmodellen door de gemeente Amsterdam'.

3.6 ***Wie zijn betrokken bij de ontwikkeling van de KI?***

3.7 ***Welke verantwoordelijkheid heeft ieder rol?***

3.8 ***Beschrijf welke actoren een rol spelen in de toepassing.***

3.9 ***Beschrijf welke KI-technologie wordt ingezet om het doel te bereiken.***

3.10 ***Beschrijf of er is nagedacht over manieren om het KI-systeem te ontwikkelen of de KI te trainen zonder of met zo weinig mogelijk gebruik van potentieel gevoelige gegevens of persoonsgegevens?***

3.11 ***Wie zorgt ervoor dat de KI operationeel (beheer) blijft en wordt gemonitord?***

4. Beschrijf de baten van de KI-toepassing

4.1 Wat zijn de baten voor de organisatie?

4.2 Wat zijn de baten voor het individu?

4.3 Wat zijn de baten voor de maatschappij als geheel?

5. Architectuur

5.1 Hoe ziet de huidige architectuur eruit voor het gebruik van de KI?

5.2 Hoe zit de toekomstige architectuur eruit voor het gebruik van de KI?

6. Data en data governance

6.1 ***Beschrijf welke brondata worden gebruikt binnen het werkproces en in het kader van de toepassing.***

6.2 ***Beschrijf of de brondata voor de KI is aangevuld met afgeleide data?***

6.3 ***Heeft de gemeentelijke organisatie volledige controle en beheersing (eigenaarschap) over de gebruikte data voor de KI?***

6.4 ***Beschrijf de reikwijdte van de gegevens in uw gegevenssets die worden onderzocht (bijv. of ze persoonsgegevens bevatten)?***

6.5 ***Wat is de bewaartermijn van de gegevens en voor verwerking ten behoeve van de KI-toepassing?***

6.6 ***Hoe is de data geleverd en verzameld?***

6.7 ***Welke afspraken zijn er gemaakt over de levering en verzameling van de data?***

6.8 ***Wat is de kwaliteit en kwantiteit van de beschikbare data?***

6.9 ***Is de data passend voor de KI-toepassing? Zo ja, waarom?***

6.10 ***De data heeft passende 'statistische' eigenschappen. Dit geldt zowel voor de individuele dataset als ook voor een combinatie van datasets***

6.11 ***Zijn er relevante aannames, met betrekking tot de informatie die de gegevens geacht worden te meten en te duiden?***

6.12 ***Zijn er 'data gaps' of tekortkoming in de data? Zo ja, kunnen er maatregelen worden troffen om dit te corrigeren? Welke maatregelen zijn dit?***

6.13 ***Zijn passende maatregelen genomen om categorieën persoonlijke data te beschermen? Zo ja, welke maatregelen zijn die (bv: encryptie, pseudonimisatie, aggregatie, anonimisatie)***

6.14 ***Welke datavoorbereiding- en verwerkingsmethodieken zijn toegepast (annotatie, labelling, cleaning, aggregatie)***

6.15 ***Zijn er afspraken gemaakt over de bewaartermijnen van de verwerkte data en de verwijdering/ archivering van de data als de termijn is verlopen of de data niet meer noodzakelijk is voor de ontwikkeling van de KI? Zo ja, wat zijn deze afspreken?***

6.16 ***Is de training-, validatie- en testdata relevant, representatief, foutloos en compleet?***

6.17 ***Zijn er duidelijke criteria/ parameters vastgesteld voor het correct functioneren van de KI?***

6.18 ***Welke methoden worden gebruikt om te verifiëren of de KI binnen de gestelde parameters blijft?***

6.19 ***Is de output van de KI consequent?***

6.20 ***Welke out-put data genereert de KI?***

6.21 ***Welke relevante designkeuzes voor de KI-toepassing zijn gemaakt?***

6.22 ***Is er onderzoek gedaan naar onbewuste vooringenomenheid in de dataset en de KI-toepassing? Zo ja, wat zijn de uitkomsten hiervan en welke maatregelen zijn getroffen?***

7. Wettige en ethische toepassing

7.1 Wat is de wettelijke grondslag voor de beoogde verwerking van de gegevens ten behoeve van het ontwikkelen, trainen en inzetten van de KI?

7.2 Welke waarden en belangen spelen een rol in de context van de KI-toepassing?
Invullen template 'weging betrokken (grond)rechten en waarden bij inzet KI

7.3 Hoe worden de waarden en belangen van betrokken actoren geborgd?

7.4 Zijn de waarden en belangen geconcretiseerd in wet- en regelgeving?
Invullen template 'weging betrokken (grond)rechten en waarden bij inzet KI

7.5 Hoe wordt gedurende het ontwikkelproces getoetst aan deze waarden en belangen?

7.6 Is er, in de gebruikssituaties waarin er mogelijk sprake kan zijn van negatieve gevolgen voor de grondrechten, een effectbeoordeling op het gebied van grondrechten uitgevoerd?

7.7 In hoeverre hebben betrokkene een kwetsbare positie ten opzichte van de gebruiker van de KI (bezien vanuit macht, kennis, economische en sociale omstandigheden of leeftijd).

7.8 Wordt door de KI bepaalde personen of groepen schade toegebracht of ongunstig behandeld wat niet gerechtvaardigd is en in verhouding staat tot hun sociale gedrag of de ernst ervan?

7.9 Wordt door keuze architectuur of de gebruikersinterface van de KI menselijk gedrag, meningen of beslissingen beïnvloed?

7.10 Versterkt of vergroot het KI-systeem menselijke capaciteiten?

7.11 Wie is de "human in control" en wat zijn de momenten of hulpmiddelen voor menselijke interventie?

7.12 In hoeverre is het mogelijk om een persoon 'uit' de KI te halen (opt-out?)

7.13 Zijn er voorzorgsmaatregelen genomen om overmatig vertrouwen in of op het KI-systeem bij werkprocessen te voorkomen?

7.14 ***In hoeverre kan de output van de KI worden teruggedraaid, gecorrigeerd of buiten beschouwing worden gelaten?***

7.15 ***Is er getoetst in hoeverre er sprake is van profilering en in hoeverre dat is toegestaan?***

7.16 ***Is er een menselijke toezichthouder aangewezen om de werking van de KI te monitoren en om effectieve maatregelen als de KI niet naar behoren presteert?***

7.17 ***Is de onderliggende logica in de KI-toepassing passend bij de wettelijke taak die moet worden uitgevoerd met de inzet hiervan?***

7.22 ***Is de KI-toepassing voldoende transparant en controleerbaar? Zo ja, waarom?***

7.23 ***Is de ontwikkelde KI-toepassing wettig en ethisch verantwoord? Zo ja, waarom?***

8. [redacted]analyse

De [redacted]analyse omvat de volgende onderdelen:

1. Te onderzoeken ("gevoelige") groepen bepalen
2. Hypotheses opstellen over features die tot indirecte [redacted] leiden
3. [redacted] selecteren passend bij het project
4. Directe [redacted] analyseren
5. Indirecte [redacted] analyseren
6. [redacted] op niet meetbare variabelen analyseren
7. Gevonden [redacted] wegen en beoordelen met verantwoordelijke directie
8. [redacted] waar nodig mitigeren
9. Conclusies opstellen

Deze onderdelen hoeven niet exact in de weergegeven volgorde uitgevoerd te worden en hebben bovendien een iteratief karakter: uitkomsten van één onderdeel kunnen ertoe leiden dat een ander onderdeel nogmaals uitgevoerd of aangepast wordt. Dit herhaalt zich totdat alle onderdelen naar tevredenheid zijn afgerond. De onderdelen worden hieronder in detail toegelicht.

8.1 Te onderzoeken ("gevoelige") groepen bepalen

Het kernprincipe wat we toepassen om [redacted] te achterhalen is om de uitkomsten van de KI te vergelijken voor verschillende groepen. De *punten* waarop wordt vergeleken (de metrics) worden beschreven in onderdeel 2. De *groepen* die worden vergeleken zijn gebaseerd op kenmerken waarop we geen ongerechtvaardigd onderscheid mogen en/of willen maken. We hebben de volgende kenmerken geïdentificeerd waarvoor dit het geval is:

- Nationaliteit*
- Geboorteland*
- Postcode*
- Huidskleur
- Etniciteit
- Geslacht*
- Leeftijd*
- Burgerlijke staat*
- Seksuele geaardheid
- Religie
- Politieke opvatting
- Verblijfsstatus*
- Zwangerschap
- Gezondheid
- Sociale klasse
- Genetica

De met een asterisk gemarkeerde kenmerken zijn bij de gemeente bekend over de mensen die, doordat zij wonen op een adres in de dataset, opgenomen zijn in de dataset.

8.2 Metrics selecteren passend bij het project

De KI kan op verschillende manieren beoordeeld worden, bijvoorbeeld:

- *Accuracy*: Percentage KI voorspellingen wat juist is.
- *Selection rate*: Percentage zaken wat door de KI aangewezen wordt als fraude.
- *False positive rate (FPR)*: Percentage van alle niet-fraudezaken dat (ten onrechte) door de KI als fraude wordt aangewezen. Zegt iets over de kans dat een zaak onnodig wordt onderzocht, daarom hoe lager, hoe beter.
- *True positive rate (TPR)*: Percentage van alle fraudezaken dat (terecht) door de KI als fraude wordt aangewezen. Zegt iets over de pakkans, daarom hoe hoger, hoe beter.
- *False negative rate (FNR)*: Percentage van alle fraudezaken dat (ten onrechte) door de KI niet als fraude wordt aangewezen. Complementair aan de TPR.
- *False discovery rate (FDR)*: Percentage van alle fraudevoorspellingen wat onjuist is.

Zulke cijfers noemen we *metrics*. Een metric kan worden uitgerekend voor de hele dataset of juist per groep, waarna we kunnen vergelijken of bepaalde groepen door de KI anders behandeld worden, waardoor zij mogelijk benadeeld worden. Het is doorgaans niet mogelijk om alle metrics tegelijkertijd te verbeteren, waardoor er met de specifieke projectcontext in gedachten keuzes gemaakt moeten worden. Zo zijn in medische toepassingen *false negatives* doorgaans ongewenst, omdat dat betekent dat er een ziekte over het hoofd wordt gezien, terwijl in andere gevallen *false positives* meer nadelen met zich meebrengen.

Om een beeld te geven van de hoeveelheid [redacted] in de KI zijn de volgende metrics gekozen:

- *Disparate impact*: Verhouding tussen de selection rate van twee groepen.
- *Average odds difference*: Gemiddelde van het verschil tussen twee groepen in FPR en het verschil in TPR.
- *False discovery rate difference*: Verschil in FDR tussen twee groepen.
- *False positive rate difference*: Verschil in FPR tussen twee groepen.
- *Generalized entropy index*: Oorspronkelijk afkomstig uit onderzoek naar inkomensongelijkheid; geeft uitdrukking aan de ongelijkheid tussen groepen én binnen een groep.

8.3 Hypotheses opstellen over features die tot indirecte [redacted] leiden

Dat een bepaald kenmerk niet als feature in de KI zit, betekent niet dat de KI hier geen [redacted] tegen kan hebben. Een bekend voorbeeld is dat KI zonder nationaliteit als feature alsnog een [redacted] op nationaliteit kan hebben doordat postcode *wel* in de KI zit en sterk samenhangt met nationaliteit. Voor alle features in de KI zijn daarom hypotheses opgesteld over mogelijke relaties met de gevoelige kenmerken. Als zo'n samenhang leidt tot een [redacted] op het gevoelige kenmerk noemen we dat indirecte [redacted]

8.4 Directe[redacted] en controlevariabelen analyseren

Met directe[redacted] wordt bedoeld een[redacted] veroorzaakt door het direct meenemen van een gevoelig kenmerk als feature in de KI. De gevoelige kenmerken die zijn overwogen voor de KI zijn geslacht en leeftijd. De exacte features die hieruit zijn afgeleid, zijn te vinden in VERWIJZING NAAR TEMPLATE

Gevoelige kenmerken die niet worden ingezet als feature kunnen worden gebruikt als controlevariabele. Dit betekent dat het gevoelige kenmerk *niet* in de KI wordt meegenomen, maar dat we er *wel* groepen mee maken om te zien of de metrics verschillen tussen deze groepen.

Het analyseren gebeurt als volgt:

1. Deel de dataset op in twee groepen op basis van een gevoelig kenmerk.
2. Bereken de bovengenoemde metrics voor beide groepen.
3. Bekijk of de metrics verschillen tussen de groepen.

Als de metrics niet (substantieel) van elkaar verschillen tussen groepen op basis van kenmerk X, dan kunnen we concluderen dat de KI geen (substantieel) verschil maakt op basis van kenmerk X. Is dit het geval, dan kunnen ook de hypotheses over indirecte[redacted] betreffende dit kenmerk van de lijst worden gehaald.

Als de metrics wel (substantieel) van elkaar verschillen, dan is er mogelijk sprake van een ongerechtvaardigde[redacted] Deze resultaten worden verzameld en meegenomen naar onderdeel 7.

8.5 Indirecte[redacted] analyseren

Met indirecte[redacted] wordt bedoeld een[redacted] veroorzaakt door het meenemen van een op het eerste gezicht onschadelijke feature die samenhangt met een gevoelig kenmerk.

Onderdeel 3 heeft geleid tot hypotheses over de features die indirecte[redacted] zouden kunnen veroorzaken. Onderdeel 4 heeft deze lijst mogelijk verkleind, als al geconcludeerd kon worden dat de KI geen verschil maakt op basis van een gevoelig kenmerk.

Het analyseren gebeurt op de manier zoals in onderdeel 4 beschreven, met als enige aanpassing dat de groepen nu niet gebaseerd zijn op een gevoelig kenmerk, maar op een van de features die voorkomt op de lijst met indirecte-[redacted]-hypotheses.

Als de metrics niet (substantieel) van elkaar verschillen tussen groepen op basis van feature X, dan kunnen we alle indirecte-[redacted]-hypotheses voor feature X ontkrachten. Als de metrics wel (substantieel) van elkaar verschillen, dan is de hypothese mogelijk waar en is er mogelijk sprake van een ongerechtvaardigde[redacted]. Deze resultaten worden verzameld en meegenomen naar onderdeel 7.

8.6 [redacted] op niet meetbare variabelen analyseren

Lang niet alle gevoelige kenmerken op de lijst in onderdeel 1 worden geregistreerd door de gemeente of zelfs de overheid in het algemeen.

8.7 Gevonden [redacted] wegen en beoordelen met verantwoordelijke directie

Het uitvoeren van de voorgaande stappen resulteert in een lijst van features met een indicatie van de mate waarin deze features [redacted] veroorzaken tegen bepaalde groepen, uitgedrukt in de metrics. Vervolgens wordt iedere feature voorgelegd aan de verantwoordelijke directie.

Wanneer een verschil in behandeling is geconstateerd (oftewel: de metrics verschillen substantieel tussen twee groepen), is de eerste vraag of dit verschil aanvaardbaar is of onaanvaardbaar. Een verschil zou bijvoorbeeld aanvaardbaar kunnen zijn als er een achterliggend verband is tussen een derde kenmerk en de uitkomsten van de KI, dat het verschil verklaart, en dit achterliggende verband is niet problematisch. Bij het nemen van deze beslissingen zijn meerdere invalshoeken relevant, hieronder weergegeven. Het is daarom van belang dat er mensen vanuit verschillende rollen en achtergronden betrokken zijn bij het wegen van de uitkomsten van de [redacted] analyse.



Wanneer een verschil aanvaardbaar wordt geacht, kan de feature behouden blijven. Wanneer een verschil niet aanvaardbaar wordt geacht, kan ervoor gekozen worden om de feature volledig uit de KI te verwijderen, of om te proberen de feature te behouden, waarover meer in onderdeel 8.

De overwegingen die gemaakt moeten worden zijn hieronder schematisch weergegeven.



8.8 [redacted] waar nodig mitigeren

De biases die niet aanvaardbaar worden geacht, worden in dit onderdeel aangepakt. De ene aanpak is het weglaten van features. De andere is om de [redacted] te proberen te verminderen door het doen van aanpassingen in de KI (bv. de loss function of postprocessing) of de data (bv. herwegen voor een gebalanceerdere dataset).

Na dit onderdeel worden de onderdelen 4 t/m 7 altijd herhaald, omdat aanpassingen aan de KI invloed kunnen hebben op de uitkomsten van de [redacted] analyse voor andere features. Bovendien is noch het weglaten van features noch het aanpassen van data of de KI een waterdichte oplossing (respectievelijk door indirecte [redacted] en imperfecties in de methodes), dus ook de [redacted] van de feature/gevoelige groep die is aangepast moet opnieuw bekeken worden.

8.9 Conclusies opstellen

Nadat alle onderdelen zijn doorlopen en in onderdeel 7 uiteindelijk alle overgebleven verschillen als aanvaardbaar beoordeeld worden, worden de conclusies van de [redacted] analyse vastgelegd in en Rapport [redacted] Analyse. Dit omvat de oorspronkelijk gevonden en uiteindelijke biases, eventuele maatregelen die genomen zijn, en de overwegingen uit onderdeel 7.

8.10 Samenvatting rapport

Nader in te vullen

9. Technische documentatie[1]

9.1 Wat zijn de interacties tussen de KI-toepassing met hard- en software dat geen onderdeel uitmaakt van de KI-toepassing zelf?

9.2 Wat zijn de revelante soft- of firmware versies?

9.3 Op welke manier is het KI-systeem in gebruik genomen en door wie?

9.4 Op welke hardware functioneert of zou de KI-toepassing moeten functioneren?

9.5 Is er een gebruikersinstructie en (indien nodig) installatie instructie opgesteld? De gebruikersinstructie bevat de volgende informatie:

- *Een gedetailleerde beschrijving van de KI-toepassing en het ontwikkelproces bestaande uit (aanvullend op KIIA);*
- *De methodiek en de uitgevoerde stappen voor het ontwikkelen van de KI-toepassing;*
- *De ontwerpspecificaties van het systeem;*
- *De belangrijkste ontwerpkeuzes inclusief afwegingen en aannames die zijn gedaan, ook met betrekking tot personen, groepen of gebruikers;*
- *De belangrijkste classificatie keuzes*
- *Waar de KI-toepassing is ontworpen en de verschillende parameters;*
- *De beslissingen over eventuele afwegingen met betrekking tot technische oplossingen die zijn gekozen om te kunnen voldoen aan de requirements;*
- *Een beschrijving van de architectuur van het systeem waaruit blijkt hoe de softwarecomponenten zijn gebouwd of zijn gekoppeld/ geïntegreerd met elkaar.*
- *De rekenmiddelen die worden gebruikt om de KI-toepassing te ontwikkelen, trainen, testen en te valideren;*

9.6 Wat zijn, indien van toepassing, de vooraf bepaalde (toekomstige) wijzigingen in de KI-toepassing en diens prestaties?

9.7 Hoe wordt omgegaan met de onvoorspelbaarheid van het handelen van de KI?

9.8 Welke maatregelen zijn genomen om de verklaarbaarheid te borgen?

9.9 Is er een overzicht van alle veranderingen die aan de KI-toepassing zijn aangebracht gedurende de levenscyclus?

[1] Zie Richtlijn European Approach for AI voor alle documentatievereisten.

9.10 Is er een beheerplan opgesteld?

10. Risicomanagement

10.1 ***Welke risicomanagement systeem wordt toegepast en welke instrumenten worden daarbij ingezet?***

10.2 ***Is er een risicoanalyse uitgevoerd over de toepassing van de KI?***

10.3 ***Is er een overzicht van de geïdentificeerde risico's?***

10.4 ***Kunnen deze risico's worden geëlimineerd of worden gereduceerd?***

10.5 ***Is er een overzicht opgesteld van de getroffen of nog te treffen maatregelen?***

10.6 ***Zijn potentiële veiligheidsrisico's vastgesteld van (andere) voorzienbare toepassingen van de technologie, met inbegrip van onbedoeld of kwaadwillig misbruik ervan? Is er een plan om deze risico's te beperken of te beheren?***

10.7 ***Is het acceptabel, gezien de risico's en de beoogde maatregelen, om de KI-toepassing in gebruik te nemen?***

10.8 ***Wordt het risicomanagementsysteem gedurende de levenscyclus van de KI-toepassing toegepast?***

10.9 ***Wordt het risicomanagementsysteem periodiek geëvalueerd?***

11. Informatiebeveiliging

11.1 Is er sprake van security by design? Zo ja, hoe wordt dit toegepast?

11.2 Is een Diepgaande Risico Analyse uitgevoerd?

11.3 Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de KI te borgen?

11.4 Welke maatregelen zijn genomen om de veiligheid van de KI te borgen?

11.5 Welke maatregelen zijn genomen om de vertrouwelijkheid te borgen?

11.6 Welke maatregelen zijn genomen om de integriteit te borgen?

11.7 Welke maatregelen zijn genomen om de beschikbaarheid te borgen?

11.8 Welke maatregelen zijn er genomen om de traceerbaarheid te borgen?

11.9 Welke maatregelen zijn genomen om de verklaarbaarheid te borgen?

11.10 Wordt loginformatie over de werking van de KI (geautomatiseerd) gegeneerd, bewaard en toegankelijk gemaakt? Deze logs bevatten minimaal:

- *Startdatum, tijd, eind data en tijd van elk gebruik*
- *De referentiedatabase waarin de data wordt gecontroleerd door de KI-toepassing*
- *De inputdata waarop de 'match' is gemaakt*
- *De indentificatie van de natuurlijke personen die betrokken zijn bij de verificatie van de resultaten.*

11.11 Door wie worden de logs bewaard?

11.12 Wordt gecontroleerd of toegangsrechten up-to-date zijn met betrekking tot de omgeving waarin de KI functioneert?

11.13 Worden toegangsrechten aangepast zodra er een uitdiensttreding of functiewijziging van een werknemer plaatsvindt?

11.14 Worden toegangsrechten uitgegeven door daarvoor bevoegde personen?

11.15 Wordt functievermenging voorkomen bij de toegang van gebruikers tot de KI?

11.16 Wordt er gebruikgemaakt van generieke beheeraccounts? Staat het aantal beheeraccounts in logische verhouding tot de beheerders?

11.17 Wordt er bij het inrichten van toegangsrechten van verschillende gebruikersgroepen/rollen gebruikgemaakt van naamgevingsconventies en systematiek?

11.18 Worden er naamgevingsconventies gebruikt voor gebruikers en beheerders, zodat zij geïdentificeerd kunnen worden?

11.19 Voeren beheerders werkzaamheden als beheerder en werkzaamheden als gewone gebruiker uit onder 2 verschillende gebruikersnamen?

11.20 Hebben gebruikersaccounts (geen) directe toegang tot onderliggende componenten?

11.21 Bestaat er een functiescheiding tussen aanvragen, autoriseren en verwerken van wijzigingen in gebruikersaccounts en toegangsrechten?

11.22 Is het wachtwoordbeheer interactief en zijn de wachtwoorden van geschikte kwaliteit?

11.23 Worden wijzigingen in de code van de KI op een gecontroleerde wijze uitgevoerd? Denk aan het testen en accorderen/autoriseren van wijzigingen

11.24 Is de KI beveiligd, zodat er geen risico is op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies?

11.25 Worden er back-ups van de KI gemaakt en kan de KI en de data hersteld worden?

12. Beheer

12.1 Is er een beheerplan opgesteld?

12.2 Is er voldoende capaciteit om de KI-toepassing operationeel te houden en wijzigingen door te voeren?

12.3 Is er voldoende kennis aanwezig om de KI-toepassing operationeel te houden en wijzigingen door te voeren?

12.4 Worden de werking van de KI-toepassing periodiek geëvalueerd?

13. Communicatie

13.1 ***Is er een gebruikersinstructie opgesteld?***

13.2 ***Zijn mechanismen ingesteld om gebruikers in kennis te stellen van de redenen en criteria achter de resultaten van het KI-systeem?***

13.3 ***Is duidelijk gemaakt richting gebruikers en betrokkenen wat het doel van het KI-systeem is en wie of wat er profijt kan hebben van het product of de dienst?***

13.4 ***Is inzichtelijk gemaakt welke kenmerken, beperkingen en potentiële tekortkomingen van het KI-systeem duidelijk kenbaar gemaakt.***

14. Verantwoording

14.1 ***Hebt u mechanismen ingesteld die de controleerbaarheid van het systeem door interne en/of externe onafhankelijke actoren mogelijk maken, bijvoorbeeld in de vorm van waarborging van de traceerbaarheid en het bijhouden van de processen en resultaten van het KI-systeem?***

14.2 ***Is er, naast de interne initiatieven of kaders voor het toezicht op ethiek en verantwoording, ook sprake van een vorm van externe begeleiding of zijn er ook controleprocessen ingesteld?***

14.3 ***Hebt u een mechanisme ingesteld om relevante belangen en waarden die door het KI-systeem in het gedrang komen, alsook potentiële afwegingen daartussen, vast te stellen?***

15. Afwegingen en conclusie

15.1 Is de toepassing proportioneel?

15.2 Is met minder ingrijpende middelen hetzelfde doel te verwezenlijken?

15.3 Gaat het om positive sum in plaats van zero sum?

15.4 Zijn er rest-risico's?

15.5 Hoe wordt verantwoordelijkheid genomen voor verder gebruik?