Samen aanjagen van vernieuwing

Praatplaat AI en Privacy

Auteur(s): SURF Privacy Expertise Centrum
Versie: 1.0
Datum: 27 maart 2024

Inhoudsopgave

1 Inleiding

Het gesprek over AI en privacy kan behoorlijk lastig zijn. Het gaat over twee complexe onderwerpen en vaak beschik je niet over de expertise op beide gebieden. Het SURF Privacy Expertise Centrum heeft daarom een praatplaat ontwikkeld om het gesprek tussen de AI expert en de privacy expert makkelijker te maken. We hebben een lijst met verkennende hulpvragen opgesteld en toegevoegd.

Dit is een eerste uitwerking waarop we graag jullie feedback ontvangen. Dus heb je aanvullingen, opmerkingen of wensen, laat het ons dan weten via pec@surf.nl. Met jullie input en feedback willen we werken aan een handreiking over AI en Privacy.

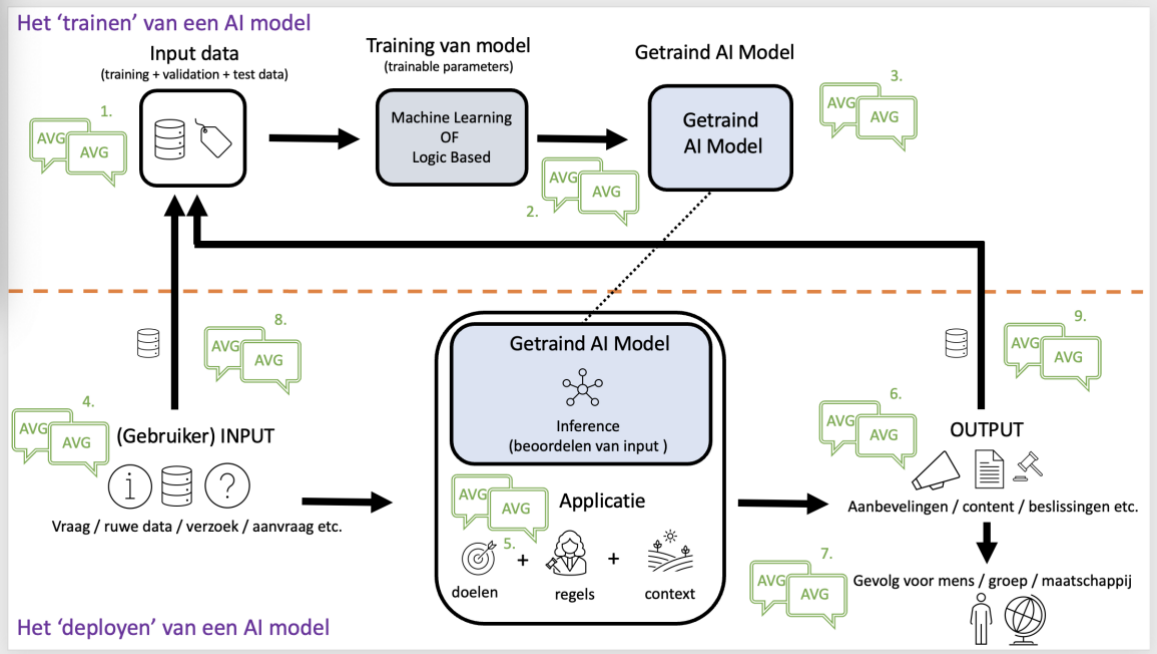
1.1 De praatplaat

De praatplaat geeft een versimpelde versie weer van:

- Het proces van het trainen van een AI model
- Het proces van het 'deployen' van een AI model
- De (mogelijke) gegevensstromen tussen deployment en hertrainen van het model.

De praatplaat heeft twee voordelen:

- Voor de AI-expert wordt het makkelijker om de privacy expert uit te leggen wat de plannen zijn en welke onderdelen van de praatplaat van toepassing zijn.
- De privacy expert kan duidelijk aangeven bij welke processen en onderdelen de AVG om de hoek komt kijken. De 'gespreksicoontjes' geven aan waar de AVG relevant is. Per 'gespreksicoontje' zijn verschillende verkennende vragen uitgewerkt.



Figuur 1: Praatplaat AI en Privacy

De tekening kan, zonder AVG icoontjes worden gebruikt voor gesprekken over andere onderwerpen dan privacy.

2 Verkennende hulpvragen over gegevensbescherming en AI

De praatplaat kan dus gebruikt worden om het gesprek tussen de AI- en de Privacy specialist de ondersteunen. In dit hoofdstuk geven we een aantal verkennende hulpvragen die aansluiten bij het gebruik van de praatplaat.

2.1 Input data (training + validation + test data) om het model te trainen

a. Waar komt de input data vandaan, bevat de data persoonsgegevens?
b. Voor welke doeleinden is de input data oorspronkelijk verzameld en past het trainen van het model binnen deze doeleinden?
c. Is het toegestaan om de data (en persoonsgegevens) te gebruiken om het model te trainen?
d. Weten personen over wie de persoonsgegevens gaan dat hun gegevens worden gebruikt voor het trainen van het model?
e. Wat is de kwaliteit van de data en is deze gecontroleerd op juistheid?
f. Hoe is gecontroleerd dat de input data geen vooroordelen en discriminerende kenmerken bevat?
g. Wat gebeurt er met de input data, nadat deze gebruikt is voor training, wordt deze bewaard? Zo ja, hoe lang?
h. Is er een duidelijk onderscheid tussen training en testdata, en indien van toepassing, worden ze gescheiden bewaard?

***AVG principes**: doelbinding, rechtmatigheid, transparantie, juistheid, behoorlijkheid, minimale gegevensverwerking*.

2.2 Training van AI model

a. Welke maatregelen zijn er getroffen om te zorgen dat het model op eerlijke wijze (ethisch, niet discriminerend) wordt getraind?
b. In hoeverre kun je reconstrueren hoe het model leert en kun je transparant zijn over diens logica (voorkomen van een black box)?

***AVG principes**: transparantie, juistheid, behoorlijkheid.*

2.3 Getraind AI model

a. Is de input data (en daarbij ook de persoonsgegevens) onderdeel geworden van het model?
b. Kan de output van het model persoonsgegevens bevatten?
c. Is het getoetst of het is toegestaan om persoonsgegevens te hebben binnen de parameters van het AI model?
d. Hoe is gecontroleerd dat het model geen vooroordelen en discriminerende kenmerken bevat?
e. Hoe is de juistheid van de output van het model geborgd?
f. Welke eisen en voorwaarden stel je aan het gebruik van het model door anderen?

***AVG principes**: rechtmatigheid, transparantie, behoorlijkheid.*

2.4 (Gebruiker) input in applicatie

a. Waar komt de (gebruiker) input vandaan, bevat de data persoonsgegevens?
b. Voor welke doeleinden en met welke grondslag is de (gebruiker) input verzameld?
c. Weten personen over wie de (gebruiker) input gaat, dat hun gegevens worden gebruikt binnen de applicatie?
d. Wat is de kwaliteit van de (gebruiker) input en is deze gecontroleerd op juistheid?
e. Is er gecontroleerd dat enkel de minimale input wordt gebruikt, die noodzakelijk is om het doel te bereiken?

***AVG principes**: rechtmatigheid, transparantie, juistheid, minimale gegevensverwerking*.

2.5 De applicatie (doelen, regels, context)

a. Wat is het doel van de applicatie?
    i. In hoeverre is het nodig persoonsgegevens te verwerken om dit doel te bereiken?
    ii. Waarom is er voor gekozen een AI model te gebruiken om het doel te bereiken?
b. Hoe gebruik je het AI model binnen de applicatie?
    i. Ken je de voorwaarden (van de ontwikkelaar van het AI model) voor het gebruik van het AI model?
    ii. Welke regels heb je zelf toegevoegd om de output te beïnvloeden?
    iii. Snap je hoe de output tot stand komt en kun je transparant zijn over diens logica (voorkomen van een black box)?
c. Hoe zorg je ervoor dat de applicatie/interface ethisch verantwoord is en er daarbij niet wordt gediscrimineerd?
    i. Biedt de applicatie passende interactievormen voor de gebruikers? In welke talen wordt het aangeboden?
d. Wat is de context waarbinnen je de applicatie gebruikt?
    i. Maakt de context de verwerking van persoonsgegevens extra gevoelig? (bijvoorbeeld medische context)
    ii. Maakt de context de inzet van AI modellen extra gevoelig? (bijvoorbeeld omdat de applicatie gericht is op kinderen)

***AVG principes:** rechtmatigheid, transparantie, juistheid, minimale gegevensverwerking.*

2.6 Output: aanbevelingen, content, beslissingen etc.

a. Bevat de output persoonsgegevens?
b. Wordt de output gecontroleerd op juistheid?
c. Wordt gecontroleerd dat de output geen vooroordelen en discriminerende kenmerken bevat?
d. Vindt er automatische besluitvorming of profiling plaats?
e. Is er sprake van menselijke tussenkomst, voordat er een besluit wordt genomen o.b.v. de output?
f. Hoe lang wordt de output bewaard?

***AVG principes**: transparantie, behoorlijkheid, rechtmatigheid, juistheid.*

2.7 Output: gevolg voor mens / groep / maatschappij

a. Wat kunnen de (negatieve) gevolgen zijn voor mens / groep / maatschappij?
    i. Is er een DPIA uitgevoerd als er waarschijnlijk hoge risico's zijn voor personen?
    ii. Zijn er maatregelen genomen om negatieve gevolgen of risico's voor personen / groepen / maatschappij te verminderen?
    iii. Worden de risico's en maatregelen periodiek geëvalueerd?
b. Hoe worden personen op de hoogte gebracht over hoe de output (+ gevolg hiervan) tot stand is gekomen?
c. Zijn er mogelijkheden om bezwaar te maken tegen de output of diens gevolgen?

***AVG principes**: transparantie, behoorlijkheid, rechtmatigheid, juistheid.*

2.8 Gebruik *(gebruiker) input* voor trainen van AI modellen

a. Komt de (gebruiker) input bij de ontwikkelaar van het AI model terecht?
    i. Zo ja, is het mogelijk om dit (technisch) te voorkomen?
        1. Zo nee: is deze verstrekking van de (gebuiker) input aan de ontwikkelaar van het AI model toegestaan?
        2. Weet je wat de ontwikkelaar van het AI model gaat doen met de (gebruiker) input?
        3. Is het gebruik van de (gebuiker) input voor verdere training van AI modellen toegestaan?
        4. Zijn de personen over wie de (gebruiker) input gaat, op de hoogte van deze verdere verwerking van diens gegevens?
b. Wordt de (gebruiker) input door de aanbieder van de applicatie voor eigen doeleinden of trainingsdoeleinden gebruikt?
    i. Zo ja, is het mogelijk om dit (technisch) te voorkomen?
        1. Is deze verstrekking van de (gebruiker) input aan de aanbieder van de applicatie toegestaan?
        2. Is de verdere verwerking van de (gebruiker) input door de applicatie aanbieder toegestaan?
        3. Zijn de personen over wie de (gebruiker) input gaat, op de hoogte van deze verdere verwerking van diens gegevens?

***AVG principes:** rechtmatigheid, transparantie, doelbinding.*

2.9 Gebruik *output* voor trainen van AI modellen

a. Wordt de output (voor trainingsdoeleinden) beoordeeld of gecorrigeerd? Zo ja, door wie?
b. Komt de output bij de ontwikkelaar van het AI model terecht?
    i. Zo ja, is het mogelijk om dit (technisch) te voorkomen? Of te anonimiseren / pseudonimiseren / minimaliseren?
        1. Zo nee: is deze verstrekking van de output aan de ontwikkelaar van het AI model toegestaan?
        2. Weet je wat de ontwikkelaar van het AI model gaat doen met de output?
        3. Is het gebruik van de output voor verdere training van AI modellen toegestaan?
        4. Zijn de personen over wie de output gaat, op de hoogte van deze verdere verwerking van diens gegevens?

c. Wordt de output door de aanbieder van de applicatie voor eigen doeleinden of trainingsdoeleinden gebruikt?
    i. Zo ja, is het mogelijk om dit (technisch) te voorkomen? Of te anonimiseren / pseudonimiseren / minimaliseren?
        1. Is deze verstrekking van de output aan de aanbieder van de applicatie toegestaan?
        2. Is de verdere verwerking van de output door de applicatie aanbieder toegestaan?
        3. Zijn de personen over wie de output gaat, op de hoogte van deze verdere verwerking van diens gegevens?

AVG principes: *rechtmatigheid, transparantie, doelbinding, minimale gegevensverwerking.*