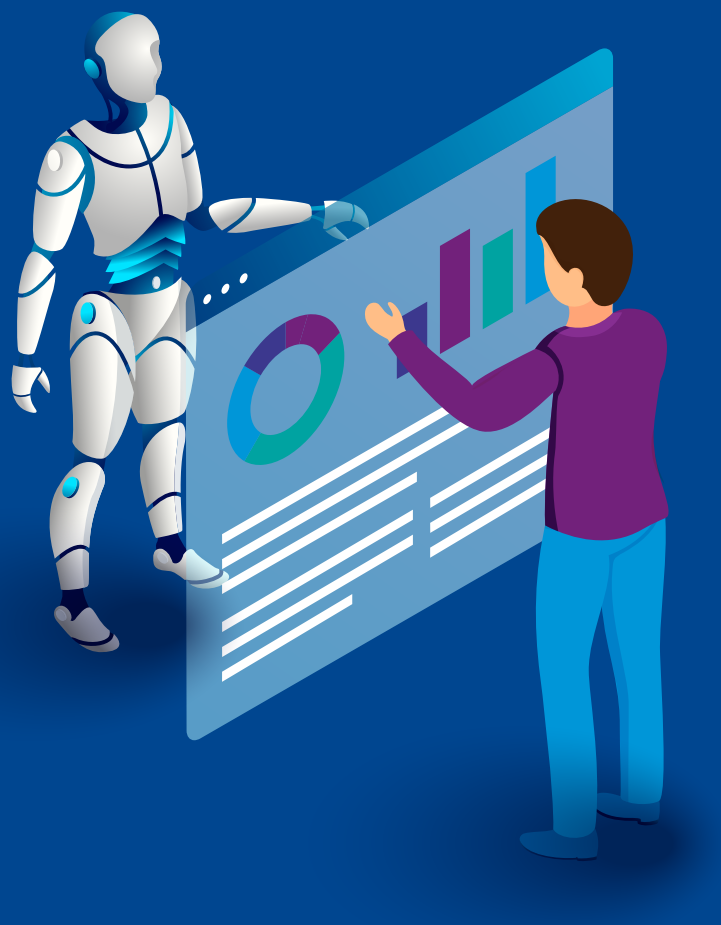
KPMG

VERBOND VAN VERZEKERAARS

# Toolkit Ethisch Kader

**Hulpmiddelen volgens KPMG bij het ethisch kader datagedreven toepassingen van het Verbond van Verzekeraars**



November 2020
Versie 1

kpmg.nl

# Inhoudsopgave

# Introductie

Het ethisch kader datagedreven besluitvorming van het Verbond van Verzekeraars (hierna: het Verbond) bevat zeven 'high-level' vereisten voor het ethisch en verantwoord gebruik van kunstmatige intelligentie (hierna: AI) en andere soorten datagedreven toepassingen door verzekeraars. Deze high-level vereisten zijn in het ethisch kader vertaald naar normen waar verzekeraars die lid zijn van het Verbond aan moeten voldoen per 1 januari 2021.

De inzet van het ethisch kader zorgt niet automatisch voor betrouwbare en ethisch verantwoorde datagedreven (/AI) toepassingen (hierna: toepassingen). Het is uw verantwoordelijkheid als verzekeraar om het ethisch kader te adopteren en hiernaar te gaan handelen. Om u hierbij op weg te helpen heeft het Verbond aan KPMG gevraagd deze toolkit op te stellen met daarin verschillende handvatten voor het operationaliseren en implementeren van de normen in het kader. **Deze handvatten vormen nadrukkelijk geen standaardaanpak en zijn daarmee volledig vrijblijvend**, en kunnen u helpen om het ethisch kader te verankeren in bestaande structuren en processen van uw organisatie.

Een pragmatische manier om met het ethisch kader aan de slag te gaan is om voor een selectie van bestaande toepassingen te beoordelen of en hoe aan het ethisch kader wordt voldaan. Zo leert u het ethisch kader en de normen goed kennen en krijgt u snel zicht op de veranderingen die in uw organisatie nodig zijn om aan het ethisch kader te kunnen voldoen. In het vervolg van dit document helpen wij u hierbij. Daarbij beginnen wij met twee modellen die u helpen met het selecteren van toepassingen die in aanmerking komen voor een eerste beoordeling. Vervolgens gaan wij per norm in op een aantal kernvragen die u kunt stellen om de geselecteerde toepassingen tegen het kader te toetsen. Waar van toepassing geven wij aanvullende informatie die van belang kan zijn om per norm mee te wegen. Onderwijl zal u merken dat er een aantal randvoorwaarden dienen te worden georganiseerd om het ethisch kader in de volle breedte te kunnen gebruiken. Ter afronding van dit document zetten wij dit voor u op een rij.

*Waar nodig zullen wij dit document in de loop van de tijd aanpassen op basis van nieuwe inzichten.*

# Het selecteren van toepassingen voor een toets tegen het kader

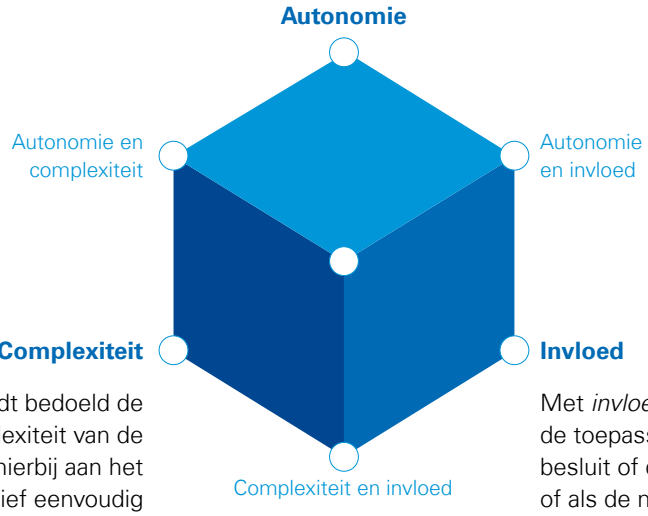
Voor een eerste beoordeling van het kader adviseren wij diverse toepassingen te selecteren. Met diverse toepassingen bedoelen wij een combinatie van de aard van de toepassing en het domein waarbinnen de toepassing wordt ingezet. Daarnaast kan diversiteit bijvoorbeeld ontstaan door verschil in complexiteit en ontwikkelstadium. Aan de hand van onderstaand classificatiemodel en domeinmodel kunt u op gestructureerde wijze een eerste selectie van bestaande toepassingen maken. Hierdoor zal u het ethisch kader via verschillende invalshoeken leren kennen.

## Classificatiemodel voor datagedreven toepassingen

Het classificatiemodel voor datagedreven (/AI) toepassingen bevat drie dimensies die helpen om de aard van een toepassing op gestructureerde wijze te bepalen. Deze drie dimensies zijn autonomie, invloed en complexiteit. Zie figuur 1 en verder voor een nadere toelichting.

Figuur 1
**Classificatiemodel voor datagedreven toepassingen**

Met *autonomie* wordt bedoeld de mate waarin een toepassing "zelf" besluiten neemt (zonder menselijke tussenkomst). Daarnaast doet autonomie zich ook voor als een aanwezige "human-in-the-loop" door gebrek aan tijd, informatie, bevoegdheden of door ondoorzichtigheid van de toepassing zelf de facto niet in staat is om de werking van de toepassing te toetsen.




Met *complexiteit* wordt bedoeld de technologische complexiteit van de toepassing. Denk hierbij aan het onderscheid tussen een relatief eenvoudig te beoordelen rekenregel/query, ten opzichte van een complex op bijvoorbeeld deep learning gebaseerd algoritme.

Met *invloed* wordt bedoeld de mate waarin de toepassing directe invloed heeft op een besluit of daaraan voorafgaande vaststelling, of als de mogelijkheid bestaat dat klanten zich door discriminatie of "singling out" harder geraakt voelen dan anderen vanwege de uitkomst(en) van de toepassing.
Invloed betekent daarmee de mate waarin een besluit voor een individu of groep rechtsgevolgen heeft, of de individu of groep op andere wijze in aanmerkelijke mate treft.

De classificatie van een toepassing wordt bepaald door per dimensie een laag, midden of hoog score toe te kennen. In de volgende tabel hebben wij een aantal karakteristieken per dimensie weergegeven die helpen om deze inschatting te maken.

Tabel 1
**Keuzehulp classificatiemodel**

| | **Laag**<br>Ethisch kader niet altijd van toepassing. | **Hoog**<br>Ethisch kader vrijwel altijd van toepassing. |
|---|---|---|
| **Autonomie** | — Er is een duidelijke "human in the loop"[1] met de tijd/ruimte om iedere uitkomst van de toepassing te beoordelen.<br>— De "human in the loop" heeft dezelfde hoeveelheid informatie ter beschikking als de toepassing.<br>— De "human in the loop" heeft het mandaat om zelf beslissingen te nemen, ook als deze tegenstrijdig zijn met de uitkomsten van de toepassing. | — Er is beperkte tijd/ruimte om de uitkomsten van de toepassing te beoordelen.<br>— Op basis van dezelfde informatie kan de "human in the loop" nooit een dergelijke beslissing nemen (bijv. door tijdgebruik of door kennisgebrek).<br>— De "human in the loop" heeft geen of zeer beperkte mogelijkheden om de uitkomst van de toepassing terzijde te leggen. |
| **Invloed** | — De uitkomsten van de toepassing hebben geen rechtsgevolgen voor een individu en/of groep.<br>— De uitkomsten van de toepassing hebben slechts in beperkte mate invloed op een uiteindelijk besluit. | — De uitkomsten van de toepassing hebben directe invloed op een individueel besluit met rechtsgevolgen of een daaraan voorafgaande vaststelling.<br>— De uitkomsten van de toepassing zorgen voor de mogelijkheid dat individuen en bedrijven zich door discriminatie of "non acceptatie" harder geraakt voelen dan anderen. |
| **Complexiteit** | — De toepassing maakt gebruik van "traditionele" data analyse technieken zoals rule based analyses.<br>— De toepassing maakt gebruik van overzichtelijke data in een vaste structuur.<br>— Kleine analyse. | — De toepassing maakt gebruik van technologie die typisch wordt aangemerkt als kunstmatige intelligentie (/machine learning).<br>— De toepassing maakt gebruik van grote hoeveelheden data, vaak ongestructureerd van aard.<br>— Er is sprake van een schakeling aan toepassingen, met ieder aannames en marges van onzekerheid.<br>— Grote analyse. |

[1] De term "human in the loop" staat voor het menselijk handelen dat rondom de toepassing plaatsvindt. Denk bijvoorbeeld aan een extra menselijke beoordeling van een uitkomst. De toepassing reikt slechts informatie aan, de "human" neemt uiteindelijk de beslissing.

*Aanvullende informatie over de complexiteit van een toepassing*

Datagedreven toepassingen is een verzamelnaam voor toepassingen die helpen beslissingen te nemen op basis van data. De technologie die hiervoor de basis vormt kan leiden tot extra risico's. In onderstaande tabel hebben wij voor de door verzekeraars meest gebruikte technologieën een aantal voorbeelden van die extra risico's per technologie weergegeven. Wij adviseren om bij het selecteren en beoordelen van de toepassingen deze risico's mee te wegen.

Tabel 2
**Meest gebruikte technologieën en risico's**

| "rule based" Toepassingen (queries, rekenregels) | Supervised machine learning | Unsupervised machine learning |
|---|---|---|
| — Ondoorzichtigheid door grote hoeveelheden regels door mensen geprogrammeerd, maar door de hoeveelheid niet meer goed te begrijpen.<br>— Onduidelijkheid over wat de gevolgen zijn van een specifieke wijziging in een deel van de toepassing op de werking van de rest van de toepassing. | — Onvoldoende "generaliseerbaarheid" naar het toepassingsdomein (ook wel overfitting[2] genoemd, al kan het zich in veel verschillende verschijningsvormen voordoen).<br>— Onjuist gebruik van metrieken (dit betekent óf de keuze, óf de interpretatie) om de accuraatheid/ prestaties van een toepassing te meten.<br>— Het gebruik van data die niet representatief is (of van onvoldoende kwaliteit) om de werkelijkheid te vertegenwoordigen. | — Mogelijkheid tot *reïdentificatie* (dit betekent het vinden van patronen in data die alsnog helpen bepaalde groepen of zelfs individuen aan te wijzen).<br>— Mogelijkheid om groepen mensen te beïnvloeden (nudging). |

## Domeinmodel voor datagedreven toepassingen

Naast de dimensies autonomie, invloed en complexiteit is ook het type dienst, product of proces waarvoor of waarin een toepassing wordt gebruikt van belang – dit wordt ook wel het toepassingsdomein genoemd. Wij adviseren om bij de selectie van toepassingen een spreiding aan te brengen in toepassingsdomeinen. Hiermee verzekert u zichzelf van een voldoende brede blik in uw eerste beoordeling van het ethisch kader. In onderstaand overzicht hebben wij een mogelijke indeling van domeinen opgenomen.

Figuur 2
**Voorbeeld domeinen voor datagedreven toepassingen**


[2] Overfitting betekent dat een getraind algoritme/model alleen bruikbaar is voor de data die is gebruikt bij diezelfde training. Wanneer het algoritme/model in "de echte wereld" wordt ingezet (generaliseerbaarheid) blijkt deze niet goed te werken.

# Leidraad voor de beoordeling van een toepassing tegen het ethisch kader

Nadat u een aantal toepassingen heeft geselecteerd kunt u starten met het per toepassing beoordelen van de normen uit het ethisch kader. Dit helpt u om inzichtelijk te maken hoe het kader in de praktijk werkt. In dit hoofdstuk hebben wij per norm een aantal voorbeeldvragen uitgewerkt. Deze vragen zullen u helpen om de beoordeling uit te voeren; bij veel normen geven wij aanvullende informatie die nuttig is om bij de beoordeling, maar ook in algemeenheid, mee te wegen.

### Normen 1 en 2

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Menselijke autonomie en controle** | Inzet van AI | 1. Voordat wij data gedreven toepassingen inzetten, voeren wij (al dan niet in het kader van een PARP proces) een adequate compliance beoordeling uit, waarbij wij een bewuste keuze maken met betrekking tot geconstateerde risico's in vergelijking tot meer traditionele technieken en processen.<br>2. Bij gebruik van data gedreven toepassingen zoals chatbots zullen wij waar nodig vermelden dat de klant met een systeem van doen heeft en niet met een mens, om verwarring of onduidelijkheid hierover te voorkomen. |

**Voorgestelde vragen**

(1) Welke risico's doen zich voor bij de inzet van de toepassing binnen het specifieke domein?
(1) Welke extra risico's doen zich voor bij de inzet van de specifieke technologie?
(1) Was/is tijdens de ontwikkeling van de toepassing voldoende domein-specifieke kennis aanwezig?
(2) Maakt de toepassing gebruik van een user interface die zichtbaar is voor de klant (bijvoorbeeld door middel van een webpagina, chatvenster, e-mail(button))?
— Zo ja, hoe ziet deze eruit?

### Normen 3 en 4

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Technische robuustheid en veiligheid** | Cybersecurity | 3. Wij zorgen dat ten aanzien van data gedreven toepassingen (inclusief databeheer) passende beveiligingsmaatregelen zijn genomen.<br>4. Wij zorgen dat data gedreven toepassingen technisch veilig en robuust zijn en alleen op basis van duidelijke kaders en onder toezicht 'zelflerend opereren'. |

**Voorgestelde vragen**

(3) Hoe ziet de (technische) architectuur (hardware, infrastructuur, server, data(base), applicatie) van de toepassing eruit?
(3) Hoe is de beveiliging van deze onderdelen georganiseerd?
(4) Is tijdens de ontwikkeling van de toepassing gebruik gemaakt van bekende (software) methodieken (ontwikkeltaal, frameworks)?
(4) Is er sprake van een zelflerend systeem? Gaat het om initiële training, hertraining of beide?
— In het geval van geautomatiseerde hertraining, hoe wordt deze gemonitord?
(4) Is er nagedacht over nieuwe beveiligingsrisico's die specifiek gelden voor kunstmatige intelligentie en machine learning.
Denk hierbij aan risico's op het gebied van *gaming the system* en *adverserial attacks*? (zie tevens aanvullende informatie)

*Aanvullende informatie*

Datagedreven toepassingen zijn vaak onderhevig aan dezelfde beveiligingsmaatregelen als reguliere ICT-toepassingen. Norm 3 zal daarom in de meeste gevallen via de standaard beveiligingsmaatregelen die gelden voor de hele organisatie zijn afgedekt, bijvoorbeeld door middel van een ISMS[3]. Echter naast bestaande beveiligingsrisico's zijn er ook nieuwe beveiligingsrisico's die ontstaan bij het gebruik van geavanceerde (data)technologie zoals kunstmatige intelligentie en machine learning. Voorbeelden hiervan zijn *gaming the system* en *adverserial attacks* waarbij buitenstaanders hun kennis over de werking van een toepassing gebruiken om bepaald gewenst gedrag van diezelfde toepassing te forceren. Het is belangrijk dat u begrijpt dat met de inzet van datagedreven toepassingen op basis van geavanceerde technologie het zogenaamde *attack surface* (de manieren waarop een aanval kan plaatsvinden) is vergroot.

[3] Een ISMS (Information Security Management System) behelst een centrale beveiligingsaanpak en is geïntroduceerd als onderdeel van de ISO27001 norm voor Informatiebeveiliging.

## Normen 5 en 6

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Technische robuustheid en veiligheid** | Fall-back en algemene veiligheid | 5. Indien een data gedreven toepassing niet (langer) technisch veilig of robuust wordt geacht, zullen wij zo spoedig mogelijk maatregelen treffen om ervoor te zorgen dat de toepassing wel voldoet. |
| | Betrouwbaarheid, reproduceerbaarheid | 6. Verzekeraars monitoren of gebruikte data gedreven systemen in overeenstemming met vooraf gestelde doelen, doelstellingen en beoogde toepassingen werken. |

**Voorgestelde vragen**

(5)(6) Wordt de toepassing gemonitord?
- — Zo ja, is er sprake van verschillende vormen van monitoring, bijvoorbeeld met verschillende frequenties (plausibiliteitscontroles, trendrapportages, etc.)?
- — In het geval van detectieve monitoring, zijn er alerts (bijvoorbeeld op basis van thresholds) ingesteld?

(5)(6) Welke opvolging is voorzien in het geval van incidenten?

*Aanvullende informatie*

Het monitoren van datagedreven toepassingen kan via verschillende methoden. In veel gevallen zal het toepassingsdomein in combinatie met de gebruikte technologie doorslaggevend zijn bij het bepalen hoe monitoring wordt ingericht. De volgende methoden, of een combinatie daarvan, zien we het meeste terug:

- — *Geautomatiseerde controles en validaties* zijn vaak integraal onderdeel van de toepassing zelf. Ze worden meestal ingezet om alle resultaten/uitkomsten van een toepassing geautomatiseerd te toetsen tegen een vooraf bepaalde verwachting. Deze verwachting kan handmatig zijn ingesteld, maar ook automatisch worden berekend op basis van eerdere uitkomsten. Op het moment dat een resultaat/uitkomst buiten bepaalde grenzen (thresholds) komt ten opzichte van de verwachting, kan er bijvoorbeeld voor worden gekozen de uitkomst niet direct door te voeren maar in een uitvalbak te plaatsen voor menselijke opvolging.
- — *Real-time monitoring* is bedoeld om direct de uitkomsten/resultaten van een datagedreven toepassing te kunnen zien en te kunnen beoordelen. Hiervoor kunnen bijvoorbeeld dashboards worden gebruikt of alerts worden ingesteld.
- — *Performance rapportages* worden met een vrij grote frequentie beoordeeld (bijvoorbeeld dagelijks) met als doel om direct inzicht te krijgen in de juiste werking (performance) van een toepassing.
- — *Trend rapportages* worden met een vrij lage frequentie beoordeeld (bijvoorbeeld maandelijks of jaarlijks) met als doel om bepaalde trends te signaleren in de uitkomsten/resultaten van een toepassing (bijvoorbeeld de Solidariteitsmonitor van het Verbond van Verzekeraars).
- — *Logging* heeft vooral als functie om in het geval van een incident of een specifiek verzoek een eerder resultaat/uitkomst te kunnen onderzoeken. Daarnaast kan een periodieke controle van de logging helpen bij het signaleren van trends (zie trend rapportages).

*Concept- en Modeldrift*

Monitoring is extra belangrijk bij datagedreven toepassingen die ontwikkeld zijn behulp van historische data (denk aan toepassingen die zijn gebaseerd op statistiek en/of machine learning). Dit zorgt namelijk voor een reële kans dat wanneer de omgeving van een dergelijke toepassing verandert, de historische data die zijn gebruikt voor de ontwikkeling ervan niet meer representatief zijn voor de nieuwe omgeving waarbinnen deze functioneert. Dit wordt ook wel *Concept drift* genoemd. Daarnaast bestaat de mogelijkheid dat door de inzet van een toepassing de verzamelde input verandert. Wanneer deze input gebruikt wordt voor een nieuwe versie van de toepassing is er sprake van *Model drift*. Voor beide vormen van *drift* is het belangrijk dat dergelijke situaties tijdig worden gesignaleerd. In veel gevallen biedt pro-actieve monitoring van de omgeving (real-time monitoring) en de uitkomsten/resultaten van de toepassing de oplossing. Op die manier kan tijdig worden gestart met het verbeteren (bijvoorbeeld door middel van hertraining) van een toepassing.

### Normen 7, 8 en 9

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Technische robuustheid en veiligheid** | Kwaliteit en integriteit van data | 7. Wij zorgen voor adequate kwaliteit (waaronder evaluatie van de datakwaliteitscriteria: volledigheid, juistheid, tijdigheid, adequaatheid en representativiteit) van (trainings)data die gebruikt wordt voor datagedreven toepassingen.<br>8. Bij gebruik van data gedreven toepassingen maken wij een weloverwogen keuze om wel of geen biometrische gegevens, data gegenereerd uit 'affective computing', social media data, online (be) zoekgegevens, locatie- en IoT-data te gebruiken en zullen klanten indien gewenst daarover transparant informeren. |
| | Toegang tot gegevens | 9. Wij zorgen voor verantwoord beheer van data en waarborging van goede data governance. |

**Voorgestelde vragen**

(7/8) Welke data zijn gebruikt om de toepassing te ontwikkelen? Wat zijn de bronnen? Wie is binnen de organisatie eigenaar van deze data en is de individu waarover de data gaat (bijvoorbeeld de klant) geïnformeerd over het gebruik van deze data?

(7/8) Hoe zijn de data die zijn gebruikt voor het ontwikkelen van de toepassing voorbereid, bewerkt, gecontroleerd en gevalideerd?

(7) In welke mate is rekening gehouden met specifieke vormen van bias inzake de data? (denk hierbij aan *survivorship bias* en *popularity bias*)[4].

(9) Welke datamanagement/-governance maatregelen heeft de organisatie getroffen?

(9) Hoe is eigenaarschap van data geregeld binnen de organisatie?

*Aanvullende informatie*

Effectief datamanagement en –governance vormt één van de belangrijkste randvoorwaarden voor het verantwoord ontwikkelen en inzetten van datagedreven toepassingen. Typisch vinden beiden plaats op centraal niveau bij een organisatie en wordt aandacht besteed aan onder andere eigenaarschap, kwaliteit, beveiliging en beschikbaarheid van data. Voor het ontwikkelen van datagedreven toepassingen is het van groot belang om te hebben nagedacht hoe data vanuit de (decentrale) bronnen/systemen worden ontsloten naar (veilige) analyse omgevingen en platformen. In het volgende hoofdstuk hebben wij een overzicht van de belangrijkste randvoorwaarden, waaronder deze, opgenomen.

[4] Survivorship en Popularity bias zijn beide veelvoorkomend. Een toelichting vindt u in deze blog: https://home.kpmg/nl/nl/blogs/home/posts/2020/04/de-onvoorspelbaarheid-van-covid-19.html

### Normen 10, 11, 12 en 13

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Privacy en data governance** | Respect voor privacy en databescherming | 10. Bij gebruik van persoonsgegevens voor data gedreven toepassingen werken wij conform de Algemene Verordening Gegevensbescherming (AVG) en de Uitvoeringswet AVG (UAVG) en de Gedragscode Verwerking Persoonsgegevens Verzekeraars.<br>11. Voorafgaand aan de inkoop, ontwikkeling en/of ingebruikname van data gedreven toepassingen voeren wij waar dit noodzakelijk blijkt een gegevensbeschermingseffectbeoordeling (een DPIA) uit.<br>12. Wij kiezen voor data gedreven systemen die zo min mogelijk potentieel gevoelige data of persoonsgegevens verwerken (dataminimalisatie) en/of waarbij de mogelijkheid bestaat om de privacy te vergroten via bijvoorbeeld encryptie, pseudonimisering/anonimisering of aggregatie.<br>13. Wij zorgen voor gedegen bescherming van (trainings)data tegen aantasting, vervuiling of hacking. |

**Voorgestelde vragen**

(10,12) Hoe is bij het gebruik van de data door de ontwikkelaar van de toepassing rekening gehouden met de eisen van de AVG (is er grondslag voor verwerking, wordt er gebruik gemaakt van bijzondere persoonsgegevens zoals gedefinieerd in de AVG?)

(11) Is er een DPIA uitgevoerd?

(12) Zijn alle data(bronnen) en specifieke variabelen die zijn gebruikt voor de ontwikkeling en de inzet van de toepassing ook strikt noodzakelijk om de toepassing optimaal te laten functioneren?

(13) Zie normen 3. en 4. en de bijpassende vragen.

*Aanvullende informatie*

De Algemene Verordening Gegevensbescherming (AVG) is op dit moment de belangrijkste geldende wetgeving voor het verzamelen en gebruiken van data. Minder bekend is dat de AVG ook een aantal belangrijke eisen stelt aan datagedreven toepassingen die worden ingezet voor geautomatiseerde besluitvorming[5]. In de tabel op de volgende pagina zijn deze eisen voor geautomatiseerde besluitvorming uit de AVG samengevat.

[5] Voor detailinformatie over geautomatiseerde besluitvorming zie artikelen 13:2f, 14:2g en 22 van de Algemene Verordening Gegevensbescherming (AVG).

Tabel 3
**Samenvatting specifieke aspecten van de AVG**

| **Aandachtspunten voor de *ontwikkeling* van datagedreven toepassingen in relatie tot het doel** | **Aandachtspunten voor de *inzet* van datagedreven toepassingen in relatie tot het doel** |
|---|---|
| *Pseudonimisatie en anonimisatie*<br>Wanneer data aantoonbaar niet meer te relateren zijn aan een individu, dan zijn het geen persoonsgegevens meer in de definitie van de AVG.<br><br>*Recht om vergeten te worden*<br>Bepaalde algoritmes kunnen ervoor zorgen dat persoonsgegevens zelf in een toepassing terecht komen (bijvoorbeeld in de vorm van een beslisregel). Dit dient voorkomen te worden.<br><br>*Bijzondere persoonsgegevens*[6]<br>— Met bepaalde technologie (bijv. unsupervised learning) is het mogelijk om bijzondere persoonsgegevens uit data op te halen, wat volgens de AVG niet is toegestaan.<br>— Wanneer algoritmes correlaties zien tussen bepaalde data die leiden tot een proxy (vervangend kenmerk) van beschermde persoonsgegevens, kan het zijn dat er alsnog verwerkingsgrondslag wordt geëist. Deze grondslag dient alsnog te worden verkregen (bijvoorbeeld door het vragen van extra toestemming).<br><br>*Transparantie over het gebruik van data*<br>De data controller (uw organisatie) dient het data subject (in veel gevallen de klant) zodanig te informeren dat eerlijke en transparante besluitvorming mogelijk wordt gemaakt. Dit kan betekenen dat er specifieke informatie over de werking van een toepassing, soms tot op het niveau van de logica, gedeeld moet kunnen worden.<br><br>*Eisen over de juistheid van data*<br>Data die worden gebruikt voor de ontwikkeling van datagedreven toepassingen dienen toereikend te zijn voor het doel dat wordt nagestreefd (dit betekent onder andere minimale gegevensverwerking).[7] | *De klant heeft altijd het recht op menselijke tussenkomst bij een besluit dat door een datagedreven toepassing wordt gegeven*<br>— In ieder geval in situaties waarbij het besluit rechtsgevolgen heeft.<br>— Het is de verantwoordelijkheid van uw organisatie om dit mogelijk te maken.<br>— De klant dient het besluit ook te kunnen betwisten en over voldoende informatie te beschikken om hier effectief van gebruik te kunnen maken.<br>— Het is de verantwoordelijkheid van de organisatie om dit mogelijk te maken.<br>— Vaak is expliciete toestemming (consent) van de klant overigens wel voldoende om geautomatiseerde besluitvorming te kunnen gebruiken.<br><br>*Recht op uitlegbaarheid*<br>De AVG verplicht uw organisatie om de klant van voldoende informatie te voorzien over de geautomatiseerde besluitvorming, zodanig dat de klant wanneer dat nodig is het besluit kan aanvechten.<br><br>*Discriminatie*<br>Het gebruik van bijzondere persoonsgegevens is toegestaan in het kader van de AVG onder strikte voorwaarden. |

**Norm 14**

| **Vereiste** | **Subvereiste** | **Norm voor verzekeraars** |
|---|---|---|
| **Privacy en data governance** | Menselijke controle | 14. Wij zorgen voor adequate training van medewerkers die met data gedreven toepassingen werken, met name ter voorkoming van 'confirmation bias' (voorkeur voor bevestiging) en met het oog op behoud van menselijke autonomie. |
| **Voorgestelde vragen**<br>(14) Zijn de medewerkers die hebben meegewerkt aan de ontwikkeling van de toepassing of er gebruik van maken voldoende getraind en ervaren om hun werkzaamheden te kunnen uitvoeren? | | |

*Aanvullende informatie*
Awareness en training vormen een belangrijke randvoorwaarde voor het op verantwoorde wijze ontwikkelen en inzetten van datagedreven toepassingen. Awareness en training kunnen plaatsvinden op centraal niveau, waarbij aandacht wordt besteed aan een grote diversiteit aan onderwerpen die verband houden met data, afhankelijk van de functie en verantwoordelijkheden van de medewerker. Ook kan er gebruik worden gemaakt van bestaande trainingen die reeds voor specifieke groepen medewerkers worden georganiseerd. In veel gevallen vindt dit decentraal plaats. In het volgende hoofdstuk hebben wij een overzicht van de belangrijkste randvoorwaarden, waaronder deze, opgenomen.

[6] Zie voor de specifieke categorieën artikel 9 van de Algemene Verordening Gegevensbescherming (AVG) en artikel 30:3b van de Uitvoeringswet Algemene Verordening Gegevensbescherming (UAVG).
[7] In artikel 5 van de AVG staat in meer detail beschreven wat wordt verstaan onder de juistheid van data.

### Norm 15

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Privacy en data governance** | Menselijk toezicht | 15. Het gebruik van data gedreven toepassingen in de praktijk vindt altijd plaats onder adequaat menselijk toezicht en menselijke verantwoordelijkheid, bijvoorbeeld door waar nodig AI te hertrainen. |

**Voorgestelde vragen**

(15) Hoe autonoom is het systeem; is er sprake van volledig geautomatiseerde besluitvorming[8]? Zo ja, welke maatregelen zijn getroffen om een "human in the loop" in te brengen?

(15) Leidt de informatie die de toepassing genereert direct tot een beslissing, of wordt die eerst nog gecombineerd met een groot aantal andere informatiebronnen?

(15) Wanneer er sprake is van menselijk toezicht; hoe is deze ingericht? Denk aan de mate waarin een betreffende medewerker:

- voldoende informatie krijgt aangereikt door de toepassing hoe een specifieke uitkomst tot stand is gekomen;
- voldoende tijd en kennis heeft om de uitkomsten van de toepassing te beoordelen;
- voldoende is uitgerust (bijvoorbeeld mandaat) om de uitkomsten van de toepassing daadwerkelijk naast zich neer te kunnen leggen.

### Norm 16

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Privacy en data governance** | Menselijk toezicht | 16. Wij zullen nieuwe technieken eerst testen in een vertrouwde setting, om te vergelijken of foutmarges en andere risico's toenemen ten opzichte van alternatieve methoden en processen. |

**Voorgestelde vragen**

(16) Is de gekozen techniek strikt noodzakelijk om de vastgestelde taak uit te voeren? Of bestaan er mogelijkheden om met een minder risicovolle techniek hetzelfde resultaat te bereiken?

*Aanvullende informatie*

Zie voor een aantal specifieke technologierisico's normen 1. en 2.

### Normen 17 en 18

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Transparantie** | n.v.t. | 17. Voordat wij data gedreven systemen inzetten bedenken wij hoe we zo goed mogelijk uitleg kunnen geven aan klanten over de uitkomsten van de toepassing.<br>18. Bij de inzet van data gedreven toepassingen zal altijd een beroep gedaan kunnen worden op menselijke tussenkomst en uitleg verkregen kunnen worden door klanten omtrent de uitkomsten bij een toepassing. |

**Voorgestelde vragen**

(17) Zie voor een aantal vragen over de gebruikersinterface norm 2.

(17) Welke specifieke maatregelen zijn er getroffen in de gebruikersinterface van de toepassing om de klant proactief te informeren over het gebruik en de werking van het systeem?

(17) In hoeverre is er in de praktijk getest/gecontroleerd hoe de uitkomsten van het systeem worden geïnterpreteerd door de belanghebbenden (klanten, medewerkers, etc.) ?

(18) Zie voor een aantal vragen over autonomie norm 15.

(18) Welke kanalen worden aan de klant aangereikt om uitleg te vragen (en krijgen) over de werking van de toepassing zelf (mits dat in relatie tot het doel van de toepassing mogelijk is, gevoelige informatie is vanzelfsprekend uitgesloten) en de uitkomst van de toepassing voor de klant?

[8] De AVG maakt een onderscheid tussen volledig geautomatiseerde besluitvorming en profiling. Deze kunnen overlappen. Bij geautomatiseerde besluitvorming gaat het steeds om besluitvorming met juridische gevolgen. Zie voor meer informatie: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling.

### Normen 19 en 20

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Diversiteit, non-discriminatie en rechtvaardigheid** | Voorkomen onrechtvaardige bias | 19. Wanneer inbreuk op grondrechten, waaronder ongerechtvaardigde discriminatoire bias in data gedreven toepassingen niet vermeden of uitgesloten kan worden, zullen wij een toepassing niet inzetten. |
| | Toegankelijkheid en inclusief ontwerp | 20. Bij de keuze voor gebruik van data gedreven systemen hebben wij oog voor diversiteit en inclusiviteit, in het bijzonder voor mensen die een risico lopen op uitsluiting of benadeling vanwege bijzondere behoeften en/of een beperking. |

**Voorgestelde vragen**

(19) Is er vóór de inzet van de toepassing onderzocht (en getest) wat de potentieel negatieve gevolgen zijn voor bepaalden groepen (discriminatie)?
— In hoeverre is hierbij ook rekening gehouden met inclusiviteit (zijn bijv. minderheden zoals chronisch zieken, mensen met een visuele beperking, etc. voldoende vertegenwoordigd)?

*Aanvullende informatie*

(het voorkomen van) bias/vooringenomenheid, discriminatie en het zorgen voor inclusiviteit vormt een aparte discipline bij de ontwikkeling van datagedreven toepassingen en vraagt om specifieke kennis[9]. Zeker wanneer er sprake is van toepassingen die gebruik maken van kunstmatige intelligentie en machine learning. Op hoofdlijnen zijn drie aspecten belangrijk:

1) De definitie van "eerlijkheid" die in de specifieke context wordt gekozen

   Er bestaat geen standaarddefinitie van "eerlijkheid" voor datagedreven toepassingen. Dit betekent dat het per toepassing kan verschillen wat er met "eerlijk" wordt bedoeld. Is dat bijvoorbeeld een maximaal prijsverschil tussen bepaalde (vaak beschermde[10]) groepen, of het feit dat de ene groep een gelijke kans heeft op een bepaald besluit als de andere groep. De keuze per toepassing wat verstaan wordt onder "eerlijkheid" dient zorgvuldig te worden afgewogen en tenminste door de eigenaar van een toepassing te worden bekrachtigd. Ook zien we dat ethische commissies hierbij een rol kunnen spelen. Idealiter vindt dit plaats aan de hand van een standaardproces.

2) De methode die wordt gebruikt om "eerlijkheid" van de toepassing te meten

   Er zijn meer dan 20 verschillende methodes om de "eerlijkheid" van een toepassing te meten. De context (/ het domein) waarbinnen een toepassing wordt gebruikt is grotendeels bepalend voor welke methode wordt gebruikt. De definitie van eerlijkheid (punt 1) is daarbij vanzelfsprekend het startpunt. Ter illustratie staan in onderstaande tabel een aantal methoden, inclusief hun manier van meten (uitkomstredenering). In aanvulling op deze methoden kunt u ook gebruik maken van specifieke toolkits om zogenaamde "feature importance", ook in relatie tot eerlijkheid, te meten. Op basis hiervan kan een uitspraak worden gedaan over de mate van eerlijkheid van een toepassing. Bekende toolkits hiervoor zijn SHAP[12] en LIME[13] .

Tabel 4
**Methoden en respectievelijke manier van meten**

| Methode[11] | Uitkomstredenering |
|---|---|
| Statistical parity | Gelijke kans (tussen twee groepen) om toegekend te worden aan een positieve voorspellende waarde. |
| Predictive parity | Gelijke kans (tussen twee groepen) om daadwerkelijk toe te horen aan de positieve voorspellende waarde. |
| Equal opportunity | Gelijke "true positive" waarden tussen twee groepen. |
| Predictive equality | Gelijke "true negative" waarden tussen twee groepen. |
| Calibration | Gelijke kans (tussen twee groepen) om daadwerkelijk tot de positieve voorspellende waarde te behoren, voor iedere score. |

[9] In deze korte beschrijving wordt het begrip discriminatie (fairness) slechts op hoofdlijnen behandeld. Discriminatie met algoritmes vormt een aparte discipline die veel meer invalshoeken heeft dan wij hier hebben weergeven.
[10] Onder beschermde groepen verstaan we groepen die door één of meerdere overeenkomstige beschermde persoonskenmerken (leeftijd, geslacht, afkomst) als groep worden gezien.
[11] De genoemde methoden zijn gericht op het meten van fairness tussen groepen (*group fairness*). Dit is de meest voorkomende vorm van fairness. Om bijvoorbeeld *invididuele fairness* en *counterfactual fairness* te meten zijn deze methoden niet geschikt.
[12] https://shap.readthedocs.io/en/latest/
[13] https://christophm.github.io/interpretable-ml-book/lime.html

3) De strategie die wordt gekozen om "eerlijkheid" van de toepassing te garanderen
In veel gevallen zal bij een eerste meting op eerlijkheid een toepassing discriminatoir blijken, zeker in gevallen waarbij historische data zijn gebruikt voor het trainen van een toepassing. In dit geval heeft de ontwikkelaar van de toepassing verschillende mogelijkheden om dit te corrigeren. Welke keuzes hierin gemaakt worden vraagt om een context-specifieke afweging. Op hoofdlijnen zijn er drie mogelijkheden. Bij 1) *pre-processing* worden er correcties op de input-data gemaakt. Vanzelfsprekend dienen dit wel juiste correcties te zijn en dienen deze vastgelegd (en onderbouwd) te worden. Bij 2) *in-processing* worden specifieke technieken gebruikt om de toepassing *tijdens* de ontwikkeling (dus in de regels van de toepassing zelf) te corrigeren. Bij 3) *post-processing* worden handmatige correcties uitgevoerd op de toepassing nadat de voorspellingen zijn gedaan.

**Normen 21 en 22**

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Maatschappelijk welzijn** | Sociale gevolgen | 21. Wij zullen de gevolgen van de inzet van datagedreven besluitvorming voor groepen klanten intern monitoren. |
| | Samenleving en democratie | 22. Wij streven ernaar om zo veel mogelijk klanten verzekerbaar te houden en zullen klanten die moeilijker of onverzekerbaar dreigen te worden informeren over manieren om risico's te verlagen of alternatieve manieren om risico's af te dekken.[14] |
| **Voorgestelde vragen**<br>(21) Zie voor een aantal vragen over monitoring norm 5.<br>(22) Zie voor een aantal vragen over transparantie normen 17. en 18. en voor een aantal vragen over inclusiviteit normen 19. en 20.<br>(22) In welke mate is onderzocht wat de impact van het algoritme is op de verzekerbaarheid van klanten?<br>— Is het management van uw organisatie hierin gekend? | | |

*Aanvullende informatie*
Beleid en strategie die ingaan op de inzet van datagedreven besluitvorming (en kunstmatige intelligentie) vormen een belangrijke randvoorwaarde voor het op verantwoorde wijze ontwikkelen en inzetten van datagedreven toepassingen. In dit beleid/strategie is ruimte om strategische keuzes te maken *wanneer* en *waarvoor* datagedreven toepassingen worden ingezet, bijvoorbeeld om maatschappelijk welzijn te bevorderen. In het volgende hoofdstuk hebben wij een overzicht van de belangrijkste randvoorwaarden, waaronder deze, opgenomen.

[14] Overeenkomstig met de Gedragscode Verzekeraars artikel 21. (https://www.verzekeraars.nl/media/5029/gedragscode-verzekeraars-2018.pdf)

### Normen 23, 24 en 25

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Verantwoording** | Controleerbaarheid | 23. Wij zorgen voor een intern controle- en verantwoordingsmechanisme voor het gebruik van AI systemen en de gebruikte databronnen.<br>24. Wij bevorderen de kennis van onze bestuurders en interne toezichthouders ten aanzien van data gedreven toepassingen.<br>25. Wij zorgen voor gedegen interne communicatie over het gebruik van data gedreven systemen. |

**Voorgestelde vragen**

(23) Hoe is vastgelegd voor de toepassing, a) wat het doel is, b) wie de eigenaar is, c) welke databronnen zijn gebruikt voor de ontwikkeling van de toepassing, d) welke databronnen worden gebruikt voor de inzet van de toepassing, e) wat de relevante risico's zijn bij de inzet van de toepassing?
— Is er een register met datagedreven toepassingen?

(24) Zie voor een aantal vragen over awareness en training norm 14. Belangrijk hierbij is om ook specifieke aandacht te geven aan het bestuur en toezichthoudende functies (bijv. 2e lijn risicomanagement, internal audit) binnen de organisatie.

(25) Zie normen 23. en 24.

*Aanvullende informatie*

Het bijhouden van een register met datagedreven toepassingen en hun karakteristieken is een belangrijke randvoorwaarde voor het op verantwoorde wijze ontwikkelen en inzetten van datagedreven toepassingen. Er kan bijvoorbeeld voor worden gekozen om aan te sluiten op het reeds bestaande verwerkingenregister in het kader van de AVG. Ook is het mogelijk om delen van het register publiek toegankelijk te maken. In het volgende hoofdstuk hebben wij een overzicht van de belangrijkste randvoorwaarden, waaronder deze, opgenomen.

Daarnaast laat de praktijk zien dat veel verzekeraars reeds gebruik maken van een manier om modellen (een vorm van datagedreven toepassingen) te classificeren. Een voorbeeld hiervan is een zogenaamde tiering-approach, waarbij modellen die onderhevig zijn aan extern toezicht vaak als tier 1 (meest risicovol) worden aangemerkt. Wij adviseren om zoveel mogelijk op deze bestaande tiering aan te sluiten, met inachtneming van de extra kenmerken die in deze toolkit bij norm 23 zijn gepresenteerd en die datagedreven besluitvorming (/AI) applicaties uniek maken.

### Normen 26, 27 en 28

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Verantwoording** | Minimalisering en verslaglegging negatieve gevolgen | 26. Wij voeren voor alle data gedreven toepassingen een risico- en effectbeoordeling uit ten aanzien van direct belanghebbenden.<br>27. Wij bevorderen de deskundigheid van onze medewerkers die werkzaam zijn op het gebied van verantwoording en controle van data gedreven systemen via een programma van educatie.<br>28. Wij zorgen voor een open cultuur in ons bedrijf waarin medewerkers worden aangemoedigd om ethische afwegingen te maken en een gedegen systeem waarbij (potentiele) negatieve gevolgen van het gebruik van een data gedreven toepassing kunnen worden gemeld en adequaat worden afgehandeld. |

**Voorgestelde vragen**

(26) Is er tijdens de ontwikkeling, en vóór de inzet van de toepassing een risicoanalyse uitgevoerd? Is deze analyse volledig en welke eventueel mitigerende maatregelen zijn getroffen?
— In hoeverre zijn de geïdentificeerde maatregelen ook aantoonbaar uitgevoerd?
(26) In hoeverre is de tweedelijnsfunctie betrokken geweest bij het maken van de risicoanalyses?
(27) Zie voor een aantal vragen over awareness en training norm 14.
(28) In hoeverre wordt medewerkers die zich in algemeenheid bezig houden met datagedreven toepassingen de mogelijkheid geboden om vragen te stellen over de afwegingen die zij moeten maken (bijvoorbeeld aan een ethische commissie, of aan een ander dergelijk orgaan)?
(28) Zie voor een aantal vragen over awareness en training norm 14.

*Aanvullende informatie*

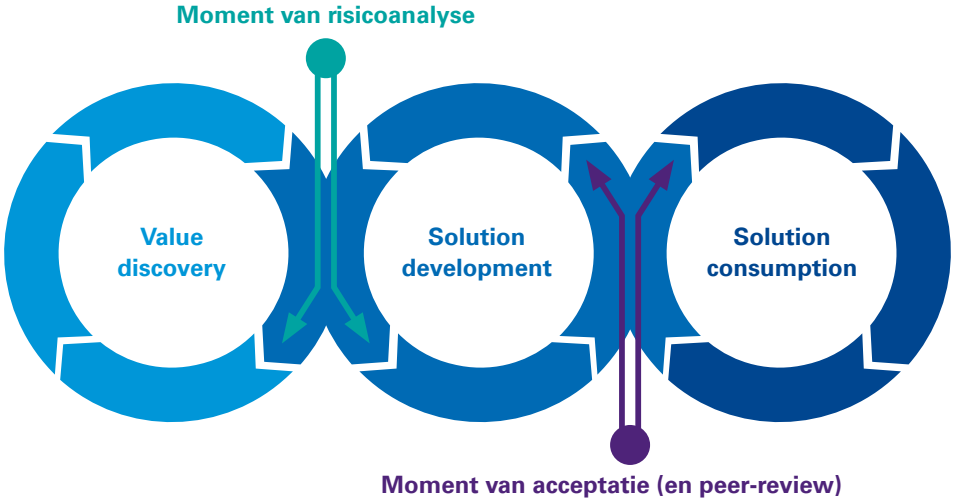
Wij verwachten dat verzekeraars die (statistische) modellen gebruiken voor bijvoorbeeld pricing en risico-inschattingen reeds maatregelen hebben getroffen om de risico's die hiermee samenhangen te beheersen: "Model Risk Management" (MRM). Typisch bestaat MRM uit een aantal bouwblokken, te weten: a) model ontwikkeling, implementatie en gebruik, b) model validatie en c) model governance. Wanneer we kijken naar datagedreven besluitvorming, zeker wanneer er gebruik wordt gemaakt van geavanceerde technieken zoals kunstmatige intelligentie en machine learning, vraagt dit om een uitbreiding van MRM. In de volgende paragrafen hebben wij per MRM-bouwblok een aantal concrete uitbreidingen voorgesteld.

*Model ontwikkeling, implementatie en gebruik*

Wij zien bij verschillende organisaties dat datagedreven (/AI) toepassingen vaak op een flexibele wijze (ook wel "agile" genoemd) worden ontwikkeld. Dit leidt tot veel pluriformiteit in hoe organisaties exact werken en daarom is het lastig om dit naar één standaard manier van werken te vertalen. Om hier enige structuur in aan te brengen hebben wij een denkkader ontwikkeld dat enerzijds helpt om meer grip te krijgen op het stadium waarin een toepassing zich in het ontwikkelproces bevindt, en anderzijds kan bijdragen aan een optimale beheersingsaanpak. Dit houdt in dat de mate van beheersing aansluit op het betreffende ontwikkelstadium (hoe meer exploratief de handelingen, hoe minder beheersing nodig is, en vice versa).

*Datagedreven toepassingen worden ontwikkeld in drie fasen*

Op basis van onze ervaring die wij hebben opgedaan bij verschillende organisaties, zien wij dat datagedreven (/AI) toepassingen vaak over de as van drie fasen worden ontwikkeld, waarbij per fase een op maat gemaakt beheersingsregime kan worden gekozen.




Tabel 5
**Toelichting fase en respectievelijke vorm van beheersing**

| Fase en toelichting | Vorm (/mate) van beheersing |
|---|---|
| **Value discovery:** De value discovery fase, waarin vrijblijvend geëxperimenteerd wordt zonder oog op operationele inzet van een toepassing. | In deze fase worden basisvoorzorgen omtrent privacy (AVG compliant) en beveiliging gewaarborgd, maar geen maatregelen specifiek voor één toepassing ingericht. Wel vindt er, voordat de overgang naar de volgende fase plaats kan vinden (het moment van opdrachtverstrekking), een risicoanalyse plaats en worden op basis van deze analyse maatregelen gekozen die tijdens de ontwikkelfase ("solution development") en de productiefase ("solution consumption") moeten worden gewaarborgd. |
| **Solution development:** De solution development fase waarin een toepassing wordt (her)ontwikkeld met het oog op operationele inzet. | In deze fase wordt een deel van de maatregelen die tijdens de risicoanalyse zijn geïdentificeerd uitgevoerd en een aantal maatregelen worden voorbereid die betrekking hebben op de volgende fase. Ook vindt, voordat overgang naar de volgende fase plaats kan vinden (het moment van acceptatie), een grondige toetsing plaats van documentatie en de juiste werking van de toepassing (mede in het kader van PARP). |
| **Solution consumption:** De solution consumption fase waarin een toepassing operationeel ingezet wordt. | Afhankelijk van het type toepassing is ook in de productiefase de mate van beheersing groot. Documentatie die betrekking heeft op voorgaande fases is in deze fase continu toetsbaar, de toepassing wordt periodiek geëvalueerd en wijzigingen die de implementatie van maatregelen raken worden gedocumenteerd. |

*Van value discovery naar solution development: het maken van een risicoanalyse (onder meer tegen het ethisch kader)*

De exploratiefase leent zich bij uitstek voor het doen van een eerste risicoanalyse met als doel te bepalen welke potentiele risico's de juiste werking van een datagedreven toepassing in de weg kunnen zitten. De risicoanalyse dient daarbij ook rekening te houden met het ethisch kader, bijvoorbeeld door per norm af te wegen of deze wel of niet van toepassing is. In het volgende overzicht hebben wij een voorbeeld opgenomen van een standaard "risico scorecard" die helpt een dergelijke analyse uit te voeren.

Door iedere toepassing aan het einde van de value discovery fase langs deze risico scorecard te leggen kan een risicoprofiel worden bepaald. Dit risicoprofiel kan vervolgens worden gebruikt als startpunt om de mate van beheersing in de tweede fase, de ontwikkelfase, vorm te geven. Door gebruik te maken van een standaard "repository" aan maatregelen (controls) kan de betrouwbaarheid van de toepassingen worden gewaarborgd[15] . Dit ziet er bijvoorbeeld als volgt uit:

| Solution Risk Scorecard | n/a | L | M | H | Tasks/ controls |
|---|---|---|---|---|---|
| The decision made by the AI application significantly affects the interests or legal position of people. | | | | V | |
| The AI application has a significant influence on the direction of attention or the information position of people. | | | | | |
| The AI application handles private, legally protected, or otherwise sensitive information about people. | | | | | |
| People are not able to make the same prediction or decision using the same information, or at least not timely and with the same frequency. | | | | | |
| Unfairness extends specifically to a subpopulation defined by a legally protected attribute (like ethnicity, gender, religion, etc.) that should be protected in that task environment. | | | | V | |
| The risk event causes significant reputation damage. | | | | V | |
| Causing the risk event is against the law, and may be interpreted as a wrongful harm deserving civil remedy or lead to enforcement by authorities. | | | | | |
| The AI application handles large amounts of money, or involves significant financial exposure. | | | | | |
| The AI application has a significant influence on markets. | | | | | |
| The AI application takes decisions fully autonomously, without supervision by people. | | | | V | |

| Example activities/controls |
|---|
| Explainability requirements<br>Explainability is integrated into the AI solution's requirements and included as one of the functional requirements of the overall solution. The requirements to the extent of explainability and type of explanations are drafted in consultation with stakeholders, with due consideration of existing legal or ethical constraints. |
| Objection procedures<br>A procedure is in place to object to the outcome of the model. Procedures for objection and human intervention are followed and steps are documented for each individual case. |
| Ethics guidelines<br>Possible ethical concerns with regards to the application are identified, it is stated how these compare to the ethical policies of the company, risks are identified and taken into consideration/mitigated during design and development. |
| Redress and compensation procedure<br>A procedure for redress and compensation of the consequences of errors is implemented. For each error found there is a documented (automatic) assessment of consequences of the error and whether redress is required (legally or otherwise). Redress procedure is followed for each of the cases where redress is deemed required. |
| Error detection mechanisms<br>An error detection system is put in place.<br>— Errors are documented<br>— documented errors are analyzed to investigate possible systematic errors |

[15] Lees hierover meer in het paper "Beheers uw algoritmes (2020)" dat beschikbaar is via https://home.kpmg/nl/nl/home/campaigns/2020/08/beheers-uw-algoritmes.html

*Het belang van validatie (van solution development naar consumption)*

Voor het valideren van datagedreven toepassingen, in welke vorm dan ook, kan gebruik gemaakt worden van bestaande methodes. Het is belangrijk dat deze methodes toezien op de doelstelling van de toepassing, de data die zijn gebruikt voor het ontwikkelen ervan, het ontwikkelproces van de toepassing en welke maatregelen zijn getroffen om de toepassing te monitoren. Wanneer er sprake is van toepassingen die gebruik maken van kunstmatige intelligentie, zijn er ten opzichte van rule-based toepassingen een aantal extra aandachtspunten van belang.

Tabel 6

**Unieke kenmerken traditionele toepassingen vs. machine learning toepassingen[16]**

| "Traditionele toepassing" | Machine learning (/AI) toepassing |
|---|---|
| Worden ontwikkeld door actuarissen met econometrische/statistische kennis die zij inzetten om aan de hand van beschikbare data toepassingen te maken. | Er is sprake van "lerend" vermogen. Op basis van input data worden inzichten min of meer automatisch in de toepassing opgenomen. De ontwikkelaar begeleidt dit proces maar bepaalt niet alles. |
| Zijn uitlegbaar. De ontwikkelaar weet precies welke aannames en keuzes gemaakt zijn bij de ontwikkeling van de toepassing. Daarnaast bepaalt de ontwikkelaar zelf welke verbanden (correlaties) in het model worden opgenomen. | Zijn, afhankelijk van het specifieke algoritme, lastig uitlegbaar. Het is niet altijd duidelijk hoe een algoritme bepaalde verbanden in de data in de toepassing heeft opgenomen. |
| Worden ontwikkeld volgens een bekende methode. Er zijn standaarden en richtlijnen beschikbaar, onder meer van de toezichthouder. | Kent veel vrijheden bij de ontwikkeling. Iedere ontwikkelaar heeft een eigen aanpak, waarbij een zeer grote verscheidenheid aan programmeertalen en "frameworks" gebruikt kunnen worden. |

Bij de methode die wordt gebruik voor validatie van een toepassing dient u met deze unieke kenmerken rekening te houden. Zo schat u de risico's beter in, met name omdat de risico's per toepassing zullen verschillen.

Het moment van validatie kan door u worden bepaald. Wij zijn van mening dat de meest grondige toetsing het beste kan plaats vinden zodra een toepassing gereed is om over te worden gedragen voor consumptie. Deze toetsing kan bijvoorbeeld een acceptatie- en/of validatietest zijn. Hierbij kan worden beoordeeld in hoeverre de maatregelen die zijn geïdentificeerd en gedocumenteerd tijdens de risicoanalyse ook zijn uitgevoerd. Ook kan hier, afhankelijk van het profiel, een keuze worden gemaakt welke diepgang wordt gehanteerd voor het uitvoeren van de review. Ook degene die de review doet (tweede lijn risicomanagement, of eventueel een "peer") kan verschillend zijn, afhankelijk van het profiel van de toepassing.

*Model governance*

Aanpassingen binnen de model governance dienen vooral gericht te zijn op het faciliteren van de uitbreidingen die hierboven bij modelontwikkeling en modelvalidatie staan beschreven.

[16] In veel organisaties wordt de term *modellen* gebruikt. Zeker als het gaat om toepassingen binnen het actuariële domein. Wij houden voor nu de term toepassingen aan, maar bedoelen hiermee nadrukkelijk ook modellen.

### Norm 29

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Verantwoording** | Documentatie van afwegingen | 29. Wij zullen gemaakte keuzes met betrekking tot de inzet van datagedreven besluitvorming in ons interne beleid vastleggen, waarbij de doorslaggevende factoren die hebben geleid tot de gemaakte keuzes ook worden vastgelegd |
| **Voorgestelde vragen**<br>(29) Zie voor vragen over het vastleggen van overwegingen over specifieke toepassingen norm 23. | | |

*Aanvullende informatie*

Beleid en strategie die ingaan op de inzet van datagedreven besluitvorming (en kunstmatige intelligentie) vormen een belangrijke voorwaarde voor het op verantwoorde wijze ontwikkelen en inzetten van dergelijke toepassingen. In dit beleid/strategie is ruimte om strategische keuzes te maken wanneer en waarvoor datagedreven toepassingen worden ingezet – onder meer om maatschappelijk welzijn te bevorderen. In het volgende hoofdstuk hebben wij een overzicht van de belangrijkste randvoorwaarden, waaronder deze, opgenomen.

### Norm 30

| Vereiste | Subvereiste | Norm voor verzekeraars |
|---|---|---|
| **Verantwoording** | Klachten | 30. Wij informeren klanten over de mogelijkheden om klachten in verband met het gebruik van datagedreven toepassingen kenbaar te maken, allereerst bij het eigen bedrijf en daarna bij aangewezen geschilbeslechters. |
| **Voorgestelde vragen**<br>(30) Zie voor vragen over transparantie en uitlegbaarheid normen 17. en 18.<br>(30) In hoeverre zijn er maatregelen getroffen om de klant in staat te stellen in beroep te gaan tegen een beslissing? | | |

# Organisatie-brede randvoorwaarden

In het vorige hoofdstuk hebben wij een aantal keer gerefereerd naar randvoorwaarden die organisatie-breed (vaak ook "centraal niveau" genoemd) moeten worden geïmplementeerd voordat het ethisch kader volledig nageleefd kan worden. Om het belang van deze randvoorwaarden te benadrukken, wordt onder meer in het verkennend rapport[17] van DNB en de AFM de Solvency-II wetgeving aangehaald die ook relevant is voor het gebruik van datagedreven toepassingen. Daarover wordt gezegd dat de organisatie verantwoordelijkheden dient te beleggen bij competente personen, maar ook dat een heldere vastlegging van het risicomanagement plaatsvindt. Bij uitstek zijn dit eisen die organisatie-brede regie vragen.

## Nieuwe organisatorische randvoorwaarden

De randvoorwaarden waar wij bij de individuele normen aan hebben gerefereerd zijn in het volgende schema als roadmap uitgewerkt. Wij maken een bewust onderscheid tussen nieuwe voorwaarden die door de introductie van het ethisch kader hoogstwaarschijnlijk voor het eerst geïmplementeerd dienen te worden en bestaande voorwaarden, die naar verwachting al (tenminste voor een deel) zijn geïmplementeerd.

Figuur 3
**Roadmap naar beheersbare datagedreven toepassingen**

### Ontwikkel een beleid voor datagedreven toepassingen

Een beleid voor datagedreven toepassingen zorgt voor richting en sturing daar waar het aankomt op het gebruik van toepassingen die zijn gebaseerd op data. In het beleid kan onder meer het volgende worden vastgelegd:
— Wat is uw organisatie's definitie van datagedreven toepassingen? (en welke technologieën vallen hieronder)
— Waarvoor mogen datagedreven toepassingen worden ingezet? En waarvoor ook niet?
— Welke positie neemt het ethisch kader in bij deze toepassingen?
Daarnaast bent u natuurlijk vrij om het beleid zo in te vullen dat het past bij uw organisatie.

### Stel een register op voor datagedreven toepassingen

Een register voor datagedreven toepassingen is een hulpmiddel om een overzicht bij te houden van alle toepassingen die binnen uw organisatie worden ingezet. In dit register legt u per toepassing verschillende gegevens vast. Denk daarbij aan het doel van de toepassing, de eigenaar, welke data worden gebruikt, wat het risicoprofiel is, enzovoort.

### Richt risicomanagement in voor datagedreven toepassingen of breid bestaande structuren uit

Door uw bestaande (model)risicomanagement tenminste uit te breiden met de aspecten die bij stap 1 zijn genoemd (en wellicht meer), zorgt u ervoor dat de risico's rondom het gebruik van datagedreven toepassingen worden beheerst. Door het ethisch kader hierin te integreren bent u al een goed stuk op weg.

### Doe aan training en awareness

Zorg ervoor dat zowel het management van uw organisatie als de medewerkers (zowel ontwikkelaars als gebruikers) op de hoogte zijn van de mogelijkheden maar ook de risico's van datagedreven toepassingen. Probeer daarbij een open cultuur te creëren waarbij zorgen over bepaalde toepassingen in openheid kunnen worden gedeeld.

### Verricht periodieke zelf-assessments

Door periodieke (zelf) assessments te (laten) verrichten zorgt u ervoor dat uw datagedreven toepassingen binnen uw "risk appetite" worden ingezet. Deze assessments dienen zowel op specifieke toepassingen te worden uitgevoerd, als ook op het gehele stelsel van risicomanagement rond datagedreven toepassingen.

[17] Zie "Artificiële Intelligentie in de verzekeringssector: een verkenning" (2019) https://www.afm.nl/nl-nl/nieuws/2019/jul/verkenning-ai-verzekeringssector

## Bestaande organisatorische randvoorwaarden

Naast de nieuwe randvoorwaarden verwachten wij dat de volgende randvoorwaarden al (groten)deels zijn ingevuld:

— Datamanagement en- governance
— Privacy
— Cyber security

Voorts geven wij in dit document hieraan geen verdere invulling.

## Een register met datagedreven toepassingen

Het maken van een register met daarin een overzicht van alle in gebruik genomen datagedreven toepassingen is een belangrijke voorwaarde om de inzet van dergelijke toepassingen op beheerste wijze te laten verlopen. Een register helpt namelijk om de specifieke en overkoepelende risico's van uw datagedreven toepassingen inzichtelijk te maken. In het register kunnen risico's worden geclassificeerd aan de hand van impact op reputatie en financiële en operationele impact. Op basis van de classificatie en vastgestelde risicoweging, kan vervolgens een risicoprofiel worden bepaald. Dit is op zichzelf stuurinformatie die helpt om uw aandacht te besteden aan de toepassingen die dit vereisen. De hoog risico toepassingen binnen uw organisatie worden op deze manier tastbaar en zichtbaar gemaakt.

In principe is het aan u om te besluiten welke gegevens in het register moeten worden opgenomen. Vaak zullen dit de standaardelementen zoals de naam van de toepassing, de eigenaar en het risicoprofiel zijn. In tabel 7 is een aantal generieke kenmerken waaraan u kunt denken opgenomen (de kenmerken zijn in het Engels omdat deze afkomstig zijn uit ons eigen oplossing voor een register, en dienen ter inspiratie).

Tabel 7
**Voorbeelden van elementen ten behoeve van het register met datagedreven toepassingen**

| **Vraag** | **Antwoordmogelijkheden** (multiple-choice) |
|---|---|
| What is the name of the use-case? If the use-case does not have a name which is consistently used throughout the organisation, please use a descriptive name. | n/a |
| Please provide a summary description of the use-case, written in such a way that an external party with limited background information would be able to understand its purpose. | n/a |
| List of data (sets) and their sources used in the analytics use-case | n/a |
| List all outputs of the use-case. The provided information should easily identify the type of output and its intended use and outcomes. | n/a |
| What is the development stage of the analytics use-case? | — Value Discovery/Proof of concept<br>— Development<br>— Solution Consumption/Production<br>— Retired |
| How is the analytics use-case developed? | — Developed fully in-house<br>— Developed by a third party<br>— Co-developed with a third party |
| Which department and team is primairly responsible for development of analytics use case? | n/a |
| Who is the owner of the use-case | n/a |
| Who is accountable for the decisions and steps taken based on the results of this analytics use-case? | n/a |
| Which department and team is the owner in? | n/a |
| Which departments and teams are use the analytics use case? Please list all teams. | n/a |
| What is the total number of users? | n/a |
| How much does a typical user use the use-case? | |

Vervolg tabel 7

| **Vraag** | **Antwoordmogelijkheden** (multiple-choice) |
|---|---|
| How much does a typical user use the use-case? | — Continuous use<br>— Daily<br>— Monthly<br>— Yearly<br>— This is a one-off solution |
| In what is the main subdomain of the organisation in which this use-case is applied? | — Customer Services / Client Relations<br>— Marketing, Sales or Public Relations<br>— Product Development/ R&D<br>— Product Portfolio Management<br>— Supply Chain or Manufacturing<br>— Service Operations<br>— Facilities<br>— IT Management<br>— Human Resources<br>— Procurement and contractor management<br>— (Financial) reporting, compliance<br>— Risk Management or Internal Audit<br>— Other (please specify) |
| What type of outputs or insights are driven by the use-case? | — Reporting to stakeholders<br>— Financial management and compliance<br>— Forecasting performance<br>— Optimizing processes<br>— Improving customer/client experience<br>— Managing risk<br>— Others (please specify) |
| What type of tools/platforms are used to develop the analytics use-case? | — BI tooling<br>— Excel<br>— Programming language<br>— Self-service data-analytics/data-science platform<br>— Specialized statistical tooling<br>— SQL<br>— Other |

Tenslotte dient u te besluiten of u een nieuw register gaat opbouwen of gebruik maakt van een bestaand register in uw organisatie (denk aan een verwerkingenregister in het kader van AVG, of een model inventory). In principe heeft het aansluiten op een bestaand register de voorkeur.

## Datagedreven toepassingen van derde partijen

Tenslotte is er een toenemende trend waarbij datagedreven toepassingen worden ingekocht bij derde partijen. Het spreekt voor zich dat in deze situaties de inzet van het ethisch kader nog steeds verplicht is. U zal dus voldoende informatie moeten verkrijgen over de toepassing zelf en uw leverancier om een toets langs het ethisch kader mogelijk te maken.

Een manier om dit te doen is door gebruik te maken van standaard inkoopvoorwaarden waarmee u eisen kan stellen aan uw leverancier aan bijvoorbeeld transparantie en uitlegbaarheid van de toepassing. In samenwerking met advocatenkantoor Pels Rijcken en de Gemeente Amsterdam hebben wij hiervoor een standaard set aan voorwaarden ontwikkeld. Deze zijn publiek toegankelijk en te vinden op de website van de gemeente Amsterdam[18].

[18] https://www.amsterdam.nl/wonen-leefomgeving/innovatie/de-digitale-stad/grip-op-algoritmes/

# Wat kan KPMG voor u betekenen?

Met deze toolkit hopen wij u verschillende handvatten te bieden om met het ethisch kader aan de slag te gaan. Mocht dit nog tot specifieke vragen leiden, zijn wij uiteraard graag bereid deze vrijblijvend voor u te beantwoorden. Onze contactgegevens vindt u aan het einde van dit document.

Als u verdere hulp kan gebruiken dan zijn wij er voor u. Wij kunnen tenminste het volgende voor u betekenen:

— **Assessment/quickscan van het ethisch kader:** het in kaart brengen van uw datagedreven toepassingen, om daaruit een aantal toepassingen te selecteren en te beoordelen langs het ethisch kader. Onze aanpak is gebaseerd op deze toolkit, al zullen wij op iedere norm een verdere verdieping uitvoeren. Wij kunnen gebruik maken van onze standaard oplossing: de Analytics Risk Index. Met deze oplossing brengen wij snel uw landschap met toepassingen in kaart. De uitkomsten kunt u tevens gebruiken als opmaat naar een register voor datagedreven toepassingen.

— **Assessment/quickscan van uw beheersingsomgeving voor datagedreven toepassingen.** Ter afsluiting van dit document hebt u kunnen lezen dat het belangrijk is om organisatie-brede maatregelen te treffen. Wij kunnen u helpen met het in kaart brengen welke “gaps” u heeft om tot volledige implementatie van het ethisch kader te komen. Hierbij beoordelen wij niet alleen een aantal individuele toepassingen, maar kijken wij ook naar uw bestaande richtlijnen, procedures en processen bijvoorbeeld vanuit uw huidige Model Risk Management structuur.

Neem contact op voor meer informatie.

# Vragen?

**Mocht u aan de hand van dit document of het ethisch kader nog vragen en/of opmerkingen hebben, neem dan gerust contact op met:**


## Frank van Praat

**KPMG Trusted Analytics**
T +31 (0) 6 512 06 152
vanpraat.frank@kpmg.nl

## Jos Schaffers

**Verbond van Verzekeraars**
T +31 (0) 70 333 86 59
j.schaffers@verzekeraars.nl