Sigra

Ethische richtsnoeren betrouwbare AI

Privacy & Informatieveiligheid

2024

Samen voor betere zorg & welzijn

Expertisecentrum Privacy & Informatieveiligheid

Privacy & Informatieveiligheid Bewustwordingscampagne

Digitale ethiek: balanceren tussen technologie en ethiek
Sigra organiseert jaarlijks in oktober de Privacy & Informatieveiligheid Bewustwordingscampagne. In 2024 is de campagne gericht op de **ethische kant** van het werken met nieuwe digitale zorg- en welzijnstechnologieën.

De ethische richtlijnen voor betrouwbare AI zijn relevant voor een breed scala aan zorg- en welzijnsprofessionals, waaronder managers, beleidsmakers, technisch specialisten en innovatiemanagers. Deze richtlijnen helpen ervoor te zorgen dat AI-toepassingen veilig, effectief en ethisch verantwoord zijn.

Campagnemiddelen
Voor Sigra-leden hebben we diverse middelen gecreëerd waarmee we professionals informeren en inspireren. Bekijk **Sigra.nl** voor meer middelen om in de eigen organisatie te verspreiden.

Sigra

Expertisecentrum Privacy & Informatieveiligheid

Ethische richtsnoeren

De Ethische richtsnoeren voor betrouwbare AI zijn opgesteld door de Onafhankelijke Deskundigengroep op Hoog Niveau inzake Artificiële Intelligentie (AI HLEG), die in juni 2018 door de Europese Commissie is opgericht.

Deze richtsnoeren bieden een kader voor het waarborgen van betrouwbare kunstmatige intelligentie en bestaan uit **drie essentiële componenten**.

- **Rechtmatigheid**: AI moet voldoen aan alle toepasselijke wet- en regelgeving.
- **Ethisch**: AI moet ethische beginselen en waarden respecteren.
- **Robuustheid**: AI moet zowel technisch als sociaal robuust zijn om onbedoelde schade te voorkomen, zelfs als de intenties goed zijn.

Deze richtsnoeren dienen als uitgangspunt voor het debat over ‘betrouwbare AI voor Europa’ en stimuleren wereldwijd onderzoek, reflectie en discussie over ethische AI-systemen.

Sigra

Expertisecentrum Privacy & Informatieveiligheid

Digitale ethiek in zorg en welzijn

Digitale ethiek in zorg en welzijn richt zich op de verantwoorde toepassing van technologie en digitale innovaties binnen de gezondheidszorg.

Het omvat onderwerpen als:

- Privacy
- Transparantie
- Data-ethiek
- AI-algoritmen
- De impact van technologie op patiënten en zorg- en welzijnsprofessionals

Innovatie vanuit waarden houdt rekening met ethische overwegingen bij het ontwikkelen en implementeren van digitale oplossingen in zorg en welzijn. Bijvoorbeeld: het waarborgen van privacy bij het delen van medische gegevens, het voorkomen van bias in algoritmen en het bevorderen van vertrouwen tussen patiënten en technologie.

Sigra

Kader voor betrouwbare AI

De richtsnoeren worden in drie lagen met verschillend abstractieniveau gegeven: van het meest abstract tot het meest concreet.

A. Grondslagen van betrouwbare AI

Hier worden de grondslagen van betrouwbare AI en de benadering op basis van grondrechten uiteengezet. De ethische beginselen die moeten worden gevolgd om ethische en robuuste AI te waarborgen, worden vastgesteld en beschreven.

B. Betrouwbare AI verwezenlijken

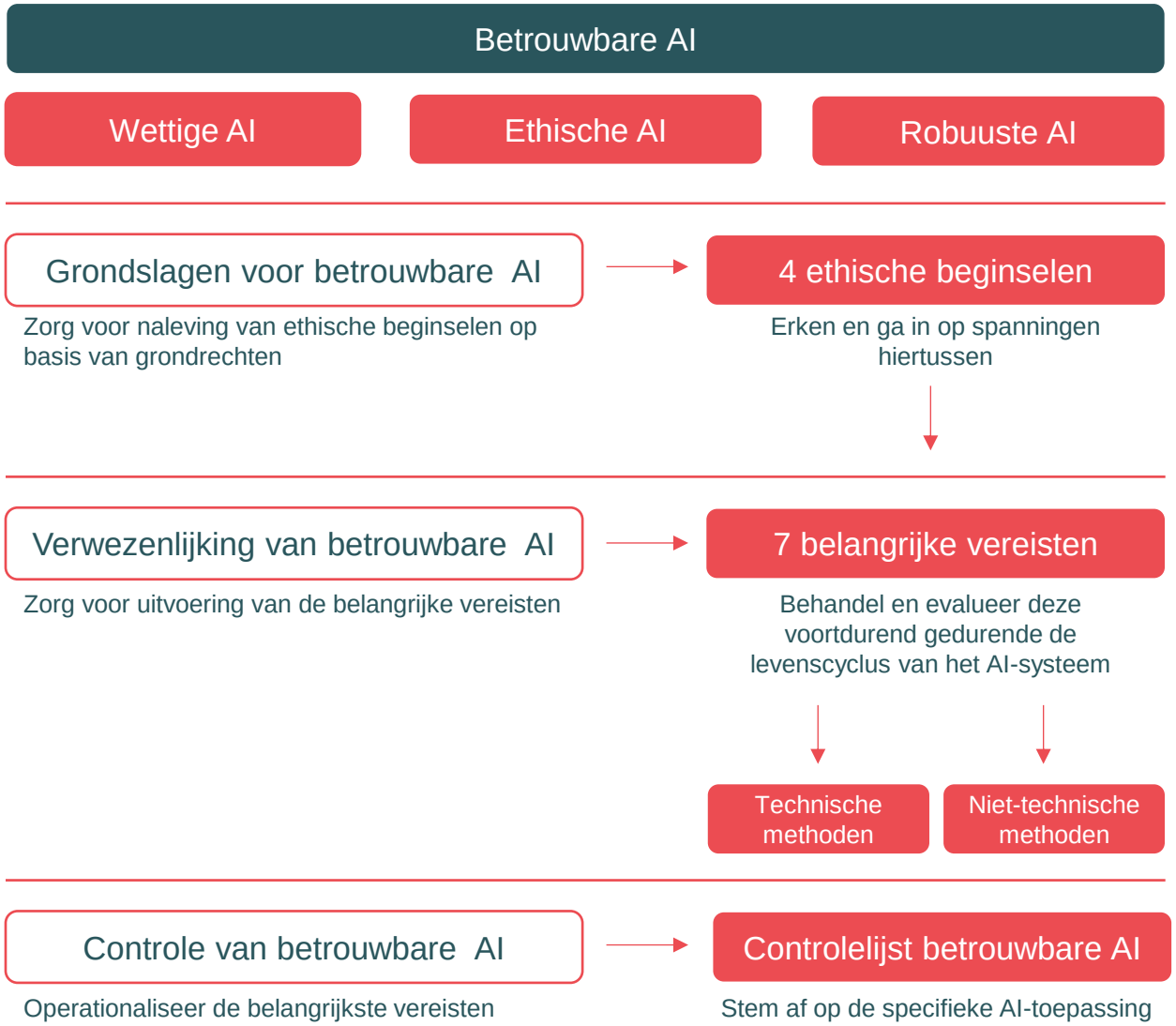
De ethische beginselen worden vertaald naar zeven vereisten die bij AI-systemen moeten worden toegepast en waaraan ze gedurende hun volledige levenscyclus moeten voldoen. Ook worden er zowel technische als niet-technische methoden aangereikt die voor de toepassing ervan kunnen worden gebruikt.

C. Betrouwbare AI controleren

Om beroepsoefenaars op weg te helpen wordt hier een voorlopige en niet-uitputtende controlelijst voor betrouwbare AI gegeven waarmee de vereisten voor betrouwbare AI kunnen worden geoperationaliseerd. Deze controle moet worden afgestemd op de toepassing van het specifieke systeem.

Expertisecentrum Privacy & Informatieveiligheid

Kader voor betrouwbare AI



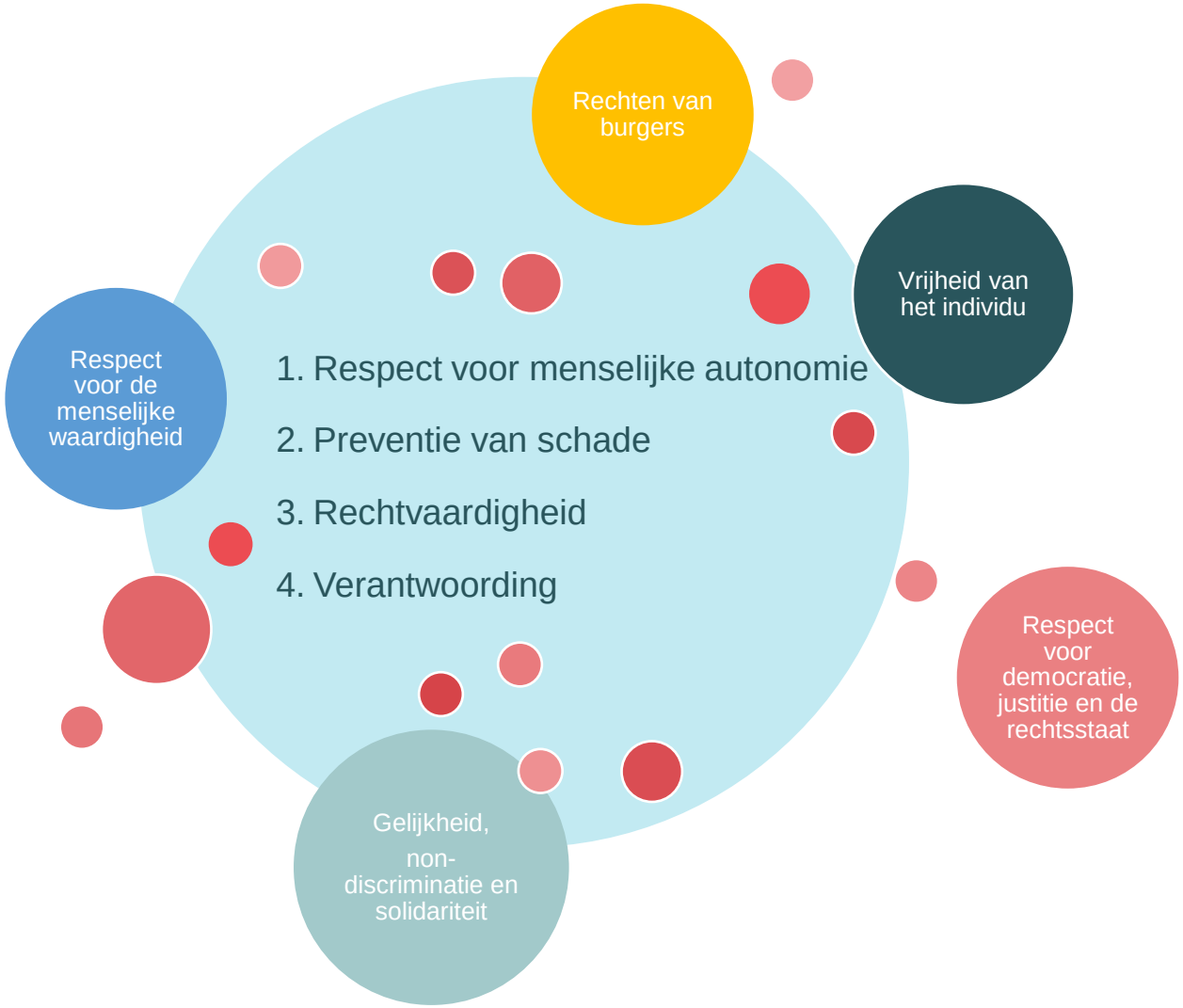
4 ethische beginselen
- Respect voor menselijke autonomie
- Preventie van schade
- Rechtmatigheid
- Verantwoording

7 belangrijke vereisten
- Menselijke controle en toezicht
- Technische robuustheid en veiligheid
- Privacy en datagovernance
- Transparantie
- Diversiteit, non-discriminatie en rechtvaardigheid
- Maatschappelijk en milieuwelzijn
- Aansprakelijkheid

Sigra

Grondslagen van betrouwbare AI

Van grondrechten naar ethische beginselen



Ethische beginselen

Respect voor menselijke autonomie

houdt in dat we respect hebben voor iemands opvattingen, keuzes en leefwijze. In de context van AI betekent dit dat we ervoor zorgen dat AI-systemen geen onterechte dwang uitoefenen op individuen. Mensen moeten autonoom kunnen beslissen over het gebruik van AI en de impact ervan op hun leven.

Kortom, dit principe waarborgt dat individuen controle houden over technologie en dat AI hun menselijke waarden en voorkeuren respecteert.

Preventie van schade

verwijst naar het waarborgen dat AI-systemen geen schade veroorzaken of verergeren voor mensen. Dit omvat bescherming van menselijke waardigheid, veiligheid, en het voorkomen van negatieve gevolgen. Het betekent ook aandacht voor kwetsbare groepen, het vermijden van ongelijkheid en het rekening houden met de natuurlijke omgeving.

Kortom, het doel is om verantwoordelijke en veilige AI te bevorderen.

Rechtmatigheid

bij AI-systemen benadrukt zowel inhoudelijke als procedurele rechtvaardigheid. Wat betreft de inhoudelijke dimensie moeten AI-systemen voordelen en kosten eerlijk verdelen, onrechtvaardige vertekening, discriminatie en stigmatisering voorkomen en gelijke kansen bevorderen. De procedurele dimensie vereist identificeerbare besluitvorming en verklaarbaarheid van het proces, zodat gebruikers effectief bezwaar kunnen maken.

Kortom, rechtvaardigheid in AI-systemen gaat niet alleen over gelijke verdeling, maar ook over transparantie en verantwoordelijkheid.

Verantwoording

bij AI-systemen benadrukt het belang van transparantie, openheid en verklaarbaarheid. Processen moeten transparant zijn, zodat gebruikers begrijpen hoe beslissingen tot stand komen. Daarnaast moeten de capaciteiten en doelen van AI-systemen duidelijk worden gecommuniceerd. In sommige gevallen, zoals bij ‘blackbox’-algoritmen, is volledige verklaring misschien niet mogelijk. Dan zijn andere maatregelen nodig, zoals traceerbaarheid en controleerbaarheid.

Kortom, verantwoording zorgt ervoor dat gebruikers vertrouwen hebben in AI-systemen door duidelijkheid te bieden over hoe beslissingen worden genomen.

Sigra

Op onze campagnepagina vind je een verdieping van de ethische beginselen.

Ethische beginselen

Respect voor menselijke autonomie

Enkele belangrijke uitdagingen:

- **Overdracht van verantwoordelijkheid** Het vinden van een evenwicht tussen menselijke en AI-verantwoordelijkheden kan lastig zijn. Te veel menselijke controle kan inefficiënt zijn, terwijl te weinig controle risico's met zich meebrengt.
- **Complexiteit van AI-systemen** Moderne AI-modellen zijn vaak complex en moeilijk te begrijpen. Het ontwerpen van transparante en interpreteerbare systemen is een uitdaging.
- **Veranderende context** AI-systemen moeten zich aanpassen aan veranderende omstandigheden. Het handhaven van menselijke controle bij dergelijke veranderingen is niet altijd eenvoudig.
- **Bias en vooroordelen** Menselijke controle kan ook leiden tot vooroordelen als de controleurs bevooroordeeld zijn. Het waarborgen van eerlijke besluitvorming is cruciaal.
- **Tijdsintensiviteit** Menselijke controle kan tijdrovend zijn, vooral bij grote datasets. Efficiënte methoden om controle uit te oefenen zijn nodig.

Sigra

Ethische beginselen

Preventie van schade

Enkele belangrijke uitdagingen:

- **Complexiteit van AI-systemen** AI-systemen kunnen complex en ondoorzichtig zijn, waardoor het voor zorgverleners moeilijk is om aan te tonen hoe schade precies is veroorzaakt. Dit bemoeilijkt het vaststellen van aansprakelijkheid.
- **Bewijslast** Bij schuldaansprakelijkheidsvorderingen moeten zorgverleners gedetailleerd aantonen wie aansprakelijk is en hoe de schade is ontstaan. Dit is lastig, vooral als AI betrokken is.
- **Lokale trainingsdata** AI-modellen in de gezondheidszorg worden vaak lokaal getraind, wat kan leiden tot beperkte generalisatie en mogelijke fouten bij bredere toepassing.
- **Acceptatie door zorgverleners** Zorgverleners moeten wennen aan AI-modellen die klinisch denken overrulen. Het gebrek aan meenemen van context en lokale factoren kan tot weerstand leiden.
- **Privacy en gegevensbescherming** De ontwikkeling van AI in de gezondheidszorg vereist betrouwbare, schaalbare en privacyvriendelijke gegevens. Het balanceren van nuttige inzichten met privacybescherming is een uitdaging.

Ethische beginselen

Rechtmatigheid

Enkele belangrijke uitdagingen:

- **Datakwaliteit en privacy** Generatieve AI-modellen worden getraind met enorme hoeveelheden data, waaronder mogelijk persoonsgegevens. Dit roept vragen op over de rechtmatigheid van het gebruik van dergelijke data.
- **Onzekerheidsmarges en transparantie** Generatieve AI-modellen geven antwoorden op basis van kansberekeningen, maar gebruikers krijgen vaak geen inzicht in onzekerheidsmarges, alternatieve antwoorden of de bronnen waarop de uitkomst is gebaseerd. Dit verhoogt het risico op verkeerde conclusies en acties.
- **Privacyrechten en intellectuele eigendom** Er zijn risico's voor privacyrechten, zoals het verkeerd koppelen van een citaat aan de verkeerde persoon. Ook kunnen generatieve systemen intellectuele eigendomsrechten schenden.
- **Vooroordelen en onevenwichtige data** Als generatieve AI-systemen worden getraind met onevenwichtige data, kunnen ze bestaande vooroordelen versterken. Dit kan leiden tot discriminatie en willekeur.
- **Concentratie van ontwikkeling bij grote techbedrijven** De ontwikkeling van generatieve foundation-modellen is momenteel geconcentreerd bij enkele grote techbedrijven, wat de machtspositie van deze bedrijven vergroot.

Sigra

Ethische beginselen

Verantwoording

Enkele belangrijke uitdagingen:

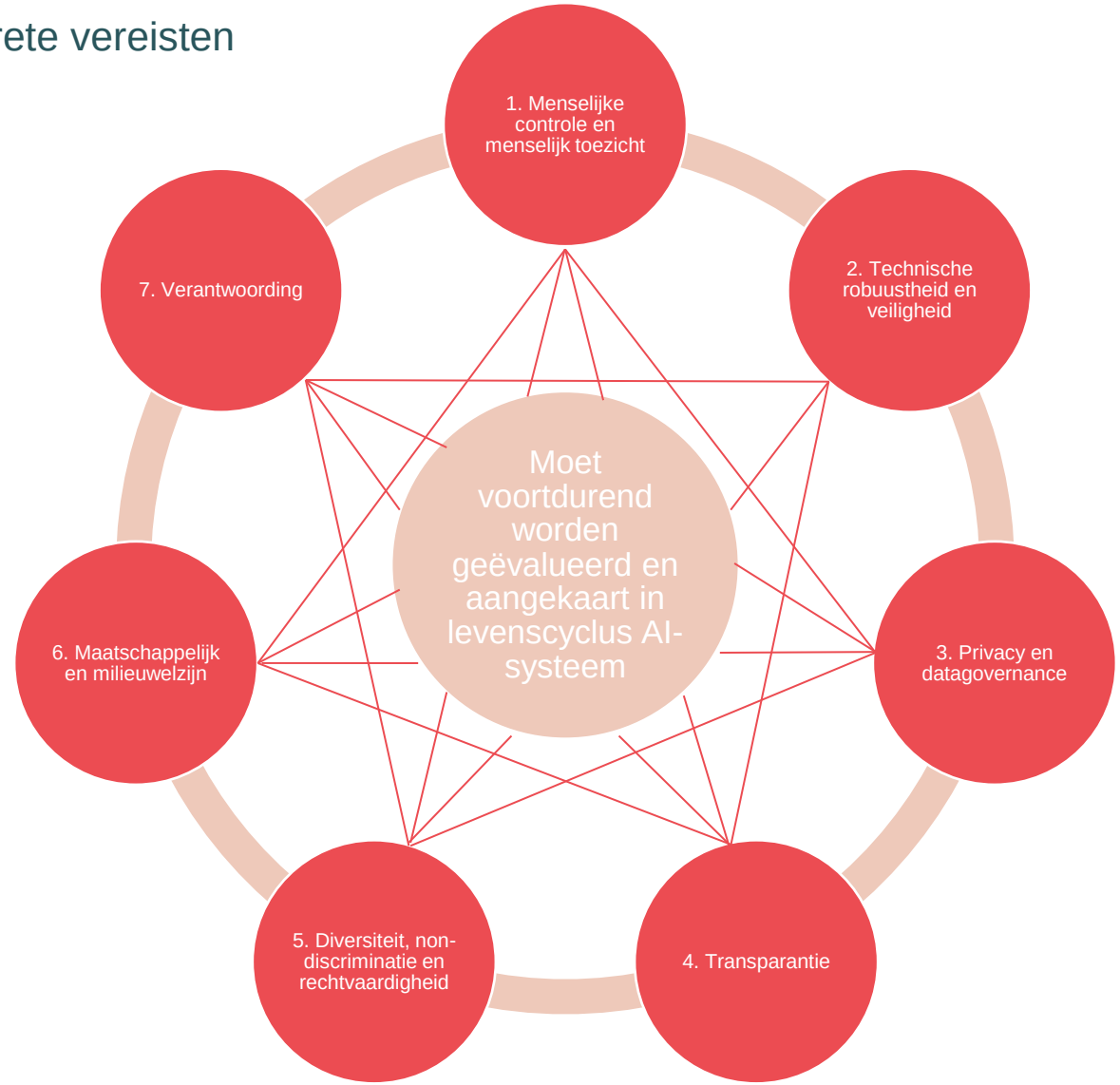
- **Gebrek aan standaardisatie** Universeel aanvaarde normen en richtlijnen ontbreken, waardoor procedures, protocollen en behandelmethoden in verschillende organisaties moeilijk te garanderen zijn.
- **Expertise en talent** Veel organisaties missen expertise en talent op het gebied van verantwoorde AI. Het ontwikkelen van verantwoorde systemen vereist gespecialiseerde kennis en vaardigheden.
- **Complexiteit van verantwoording** Verantwoording omvat meerdere componenten, zoals context, bereik, agenten, normen en processen. Het beheren van al deze aspecten kan uitdagend zijn.
- **Transparantie en uitlegbaarheid** Het realiseren van transparante en uitlegbare AI-systemen is lastig, vooral bij complexe modellen zoals “blackbox”-algoritmen.
- **Privacybescherming** Privacy is een cruciaal aspect van verantwoorde AI. Het waarborgen van privacy bij het gebruik van AI blijft een voortdurende uitdaging.

Sigra

Betrouwbare AI verwezenlijken

Vereisten van betrouwbare AI

Beginselen vertaald naar concrete vereisten



Sigra

Vereisten van betrouwbare AI

1. Menselijke controle en menselijk toezicht	Omvat grondrechten, menselijke controle en menselijk toezicht.
2. Technische robuustheid en veiligheid	Omvat weerbaarheid tegen aanvallen en beveiliging, een uitwijkplan en algemene veiligheid, nauwkeurigheid, betrouwbaarheid en reproduceerbaarheid.
3. Privacy en datagovernance	Omvat respect voor privacy, de kwaliteit en integriteit van gegevens en toegang tot gegevens.
4. Transparantie	Omvat traceerbaarheid, verklaarbaarheid en communicatie.
5. Diversiteit, non-discriminatie en rechtvaardigheid	Omvat het voorkomen van onrechtvaardige vertekening, toegankelijkheid en universeel ontwerp en participatie van belanghebbenden.
6. Maatschappelijk en milieuwelzijn	Omvat duurzaamheid en milieuvriendelijkheid, sociale gevolgen, de samenleving en de democratie.
7. Verantwoording	Omvat controleerbaarheid, minimalisering en verslaglegging van negatieve gevolgen, afwegingen en beroep.

Verwezenlijking van betrouwbare AI

Verwezenlijking van betrouwbare AI gedurende de gehele levenscyclus van het systeem.

Verwezenlijking van betrouwbare AI

Technische methoden	Niet-technische methoden
• Architecturen voor betrouwbare AI	• Regelgeving
• Ethiek en rechtsstaat door ontwerp	• Gedragscodes
• Verklaringsmethoden	• Normalisatie
• Testen en valideren	• Certificering
• Kwaliteit van de dienstindicatoren	• Verantwoording via governance kaders
	• Onderwijs en bewustzijn ter bevordering van ethische denkwijze
	• Participatie van belanghebbenden en sociale dialoog
	• Diversiteit en inclusieve ontwerpteams

Voor details zie Ethische richtsnoeren voor betrouwbare AI.

Betrouwbare AI controleren

Sigra

Expertisecentrum Privacy & Informatieveiligheid

Betrouwbare AI controleren

De richtsnoer beschrijft een niet-uitputtende controlelijst voor betrouwbare kunstmatige intelligentie. Deze lijst is bedoeld voor AI-systemen met rechtstreekse interactie met gebruikers en is gericht op ontwikkelaars en installateurs, ongeacht of zij het systeem zelf hebben ontwikkeld of van derden hebben verkregen.

De controlelijst behandelt niet de operationalisering van de eerste component van betrouwbare AI (wettige AI).

Het volgen van de lijst vormt geen bewijs van wettelijke naleving en is ook geen richtsnoer voor toepasselijke wetgeving. Aangezien AI-systemen sterk afhankelijk zijn van de specifieke context, moet de controlelijst worden afgestemd op de gebruikssituatie.

Sigra

Betrouwbare AI controleren

Governace

Het is van belang dat zorgaanbieders nadenken over de toepassing van de controlelijst voor betrouwbare AI binnen hun organisatie. Dit kan worden bereikt door het controleproces op te nemen in bestaande governance mechanismen of door nieuwe processen te introduceren.

De keuze hiervoor hangt af van de interne structuur, omvang en beschikbare middelen van de organisatie.

Controlelijst voor betrouwbare AI gebruiken

De controlelijst is bedoeld als leidraad voor beroepsbeoefenaars op het gebied van AI bij het ontwikkelen, installeren en gebruiken van betrouwbare AI. De controle moet op evenredige wijze worden afgestemd op de specifieke gebruikssituatie. Gedurende de testfase kunnen er specifieke gevoelige gebieden naar voren komen.

- Assessment List for Trustworthy Artificial Intelligence (ALTAI)
- AI Impact Assessment

Verband met bestaande wetgeving en processen

Veel beroepsbeoefenaars op het gebied van AI beschikken ook al over bestaande controlehulpmiddelen en softwareontwikkelingsprocessen om naleving, ook van niet-wettelijke normen, te waarborgen.

De onderstaande controle hoeft niet noodzakelijkerwijs los te worden uitgevoerd, maar kan in dergelijke bestaande praktijken worden verwerkt.

Bron: Onafhankelijke Deskundigengroep op Hoog Niveau inzake Artificiële Intelligentie (AI HLEG)

Sigra

Expertisecentrum Privacy & Informatieveiligheid

Meer informatie

Sigra is een regionaal samenwerkingsverband van organisaties in zorg en welzijn in Noord-Holland.

Over het expertisecentrum
Vanuit het Expertisecentrum Privacy & Informatieveiligheid helpen we leden om de privacy en informatieveiligheid in de organisatie goed te organiseren. Je kunt hier terecht voor ondersteuning en advies en om ervaringen met andere Sigra-leden uit te wisselen. Bekijk Sigra.nl voor meer informatie.

Over de campagne
Jaarlijks organiseert Sigra in oktober een Privacy & Informatieveiligheid Bewustwordingscampagne. Bekijk Sigra.nl voor meer informatie.

Contact
Het Expertisecentrum Privacy & Informatieveiligheid is te bereiken via: pi@sigra.nl.

Sigra