Rijksoverheid

Gids AI-verordening

Versie 1.0 – oktober 2024

Beeld omslag: Wes Cockx & Google DeepMind / BetterImages of AI / AI large language models / CC-BY 4.0

Leeswijzer en disclaimer
U ontwikkelt AI-systemen of wil deze misschien wel gaan toepassen in uw organisatie. Dan zult u mogelijk met de AI-verordening te maken gaan krijgen. Voor u is deze gids samengesteld, een hulpmiddel om op een toegankelijke manier inzicht in de belangrijkste aspecten van de AI-verordening te krijgen. **Aan de inhoud van deze gids kunnen geen rechten worden ontleend.** De wettekst van de AI-verordening blijft altijd leidend.

Heeft u feedback op deze gids? E-mail dan naar ai-verordening@minezk.nl. Uw feedback wordt gebruikt om deze gids te verbeteren.

Leest u een geprinte versie van deze gids? Op ondernemersplein.nl/AIA vindt u altijd de laatste versie.

De AI-verordening

De AI-verordening is een uitgebreide wet over artificiële intelligentie (AI) voor de hele Europese Unie (EU). In de AI-verordening staan regels voor het verantwoord ontwikkelen en gebruiken van AI door bedrijven, overheden en andere organisaties met als doel om de veiligheid, gezondheid en fundamentele rechten van mensen te beschermen. Hierdoor kunnen organisaties erop vertrouwen dat de AI die zij gebruiken verantwoord is en kunnen ze optimaal profiteren van de kansen van AI.

De verordening gaat in stappen in en zal medio 2027 voor het grootste deel van toepassing zijn. Een aantal AI-systemen zijn vanaf februari 2025 al verboden. Daarom is het slim om u nu alvast voor te bereiden. Om u hierbij te helpen, worden in deze gids de belangrijkste bepalingen uit de AI-verordening genoemd. Er kunnen geen rechten ontleend worden uit dit document, dat puur ter ondersteuning dient. De volledige wettekst vindt u hier.[1]

Wat betekent de AI-verordening voor uw organisatie?

Afhankelijk van het AI-systeem en wat een organisatie met dat systeem doet, gaan er eisen gelden voor de ontwikkeling en het gebruik van dit systeem. Of er eisen gaan gelden, is onder andere afhankelijk van het risico dat het AI-systeem heeft voor veiligheid, gezondheid en fundamentele rechten. Ook worden er andere eisen gesteld aan organisaties die AI (laten) ontwikkelen, dan aan organisaties die AI gebruiken. Om erachter te komen wat de AI-verordening voor uw organisatie betekent, is het belangrijk om de volgende vier stappen te doorlopen. Deze stappen worden in de rest van de gids toegelicht:

Stap 1 (Risico): *Valt ons (AI-)systeem binnen één van de risicocategorieën?*
Stap 2 (AI): *Is ons systeem 'AI' volgens de AI-verordening?*
Stap 3 (Rol): *Zijn wij de aanbieder of gebruiksverantwoordelijke van het AI-systeem?*
Stap 4 (Verplichtingen): *Aan welke verplichtingen moeten wij ons houden?*

[1] https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX:32024R1689

Stap 1. (Risico) Valt ons (AI-)systeem binnen één van de risicocategorieën?

Alle AI-systemen vallen onder de AI-verordening, maar er worden verschillende eisen gesteld aan verschillende categorieën, gebaseerd op het risico. Dit risico wordt bepaald door de beoogde **toepassing of product** waarvoor het AI-systeem wordt ontwikkeld, verkocht en gebruikt:

- **Verboden AI**: deze AI-systemen mogen niet verkocht, gebruikt of in gebruik gesteld worden.
- **Hoog-risico AI**: deze AI-systemen moeten aan verschillende eisen voldoen om risico's te beperken voordat ze mogen worden verkocht of gebruikt.

Daarnaast gaan er eisen gelden voor AI-modellen en AI-systemen die specifieke dingen **kunnen**:

- **AI-modellen en systemen voor algemene doeleinden**: hier gaan specifieke informatie-eisen voor gelden. In sommige gevallen moet ook voldaan worden aan eisen om risico's te beperken.
- **Generatieve AI en chatbots**: hier gaan specifieke transparantie-eisen voor gelden onafhankelijk of het systeem hoog-risico is of niet.[2]

Soms kan hetzelfde AI-systeem onder meerdere categorieën vallen. Zo kan een chatbot ingezet worden voor een hoog-risico toepassing. AI-systemen die niet onder één van de bovenstaande categorieën vallen, hoeven ook niet uit eisen uit de AI-verordening te voldoen. Het kan wel zo zijn dat deze systemen nog steeds aan eisen uit andere regelgeving moeten voldoen, zoals de Algemene Verordening Gegevensbescherming (AVG).

Om te weten of u aan eisen uit de AI-verordening moet voldoen, is het van belang om eerst te bepalen onder welke categorie uw AI-systeem valt. Hieronder worden de verschillende risico-categorieën uitgebreid toegelicht.

1.1. Verboden AI-systemen

Bepaalde AI-praktijken brengen een onaanvaardbaar risico met zich mee voor mensen en de samenleving en zijn daarom vanaf februari 2025 verboden. Dit betekent dat deze systemen niet in de handel gebracht, gebruikt of in gebruik gesteld mogen worden. Deze verboden gelden zowel voor **aanbieders** als **gebruiksverantwoordelijken** (verder toegelicht onder Stap 3. Zijn wij de aanbieder of gebruiksverantwoordelijke van het AI-systeem? op pagina 10).

[2] Er gaan ook specifieke transparantie-eisen gelden voor emotieherkenning- en biometrische categoriseringssystemen. Omdat dit ook hoog-risico AI-systemen zijn, zijn deze eisen opgenomen onder de hoog-risico verplichtingen in stap 4.2.

Verboden AI-systemen

1. Systemen die erop zijn gericht **menselijk gedrag te manipuleren** om de vrije keuze van personen te beperken en die tot aanzienlijke schade voor die personen kunnen leiden.
2. Systemen die **misbruik maken van de kwetsbaarheden** van personen als gevolg van hun leeftijd, handicap of specifieke sociale of economische situatie en tot aanzienlijke schade voor die personen kunnen leiden.
3. Systemen voor het opstellen van belonende en bestraffende puntensystemen gebaseerd op sociaal gedrag of persoonlijke kenmerken, bekend als **social scoring**, met een nadelige en oneerlijke behandeling tot gevolg.
4. Verbod op systemen voor **risicobeoordelingen voor het plegen van strafbare feiten**, uitsluitend op basis van profilering of (persoonlijkheids)kenmerken.
5. Systemen die **databanken voor gezichtsherkenning** aanleggen of uitbreiden door middel van het ongericht **scrapen** van gezichtsafbeeldingen van het internet of CCTV-beelden.
6. Systemen voor **emotieherkenning** in de werkomgeving en het onderwijs, tenzij dit wordt gedaan voor medische of veiligheidsredenen.
7. Systemen die worden gebruikt om **mensen op basis van biometrische gegevens te categoriseren** in bepaalde gevoelige categorieën.
8. Het **gebruik van biometrische identificatie op afstand in real time in openbare ruimten met het oog op de rechtshandhaving**. Er gelden enkele uitzonderingen in gevallen waarin het gebruik strikt noodzakelijk is, bijvoorbeeld voor het zoeken naar specifieke slachtoffers van ontvoering en mensenhandel of vermiste personen. Voor deze toepassingen gelden wel nog aanvullende waarborgen.

1.2. Hoog-risico AI-systemen

Hoog-risico AI-systemen kunnen risico's hebben voor de gezondheid, veiligheid of fundamentele rechten van personen, zoals het recht op privacy en om niet gediscrimineerd te worden. Tegelijkertijd kunnen deze systemen ook toepassingen hebben met positieve effecten op mensen en organisaties, als ze betrouwbaar zijn en de risico's worden beperkt. Daarom moeten hoog risico AI-systemen vanaf augustus 2026 aan verschillende eisen voldoen voordat ze in de handel gebracht, gebruikt of in gebruik gesteld mogen worden. Dit betekent dat **aanbieders** er tijdens de ontwikkeling van het systeem voor moeten zorgen dat het aan de eisen voldoet voordat het voor het eerst verkocht of gebruikt wordt. Een professionele partij die het AI-systeem onder eigen verantwoordelijkheid gebruikt, is een **gebruiksverantwoordelijke** (verder toegelicht onder Stap 3. Zijn wij de aanbieder of gebruiksverantwoordelijke van het AI-systeem? op pagina 10). Voor gebruiksverantwoordelijken gelden ook verplichtingen, om risico's als gevolg van de specifieke inzet van het systeem te beperken.

Er zijn twee typen hoog-risico AI-systemen:

- **Hoog-risico producten:** AI-systemen die direct of indirect ook onder een selectie van **bestaande productregelgeving** (zie hieronder) valt. Bijvoorbeeld een AI-systeem als veiligheidsonderdeel van een lift of een AI-systeem dat een medisch hulpmiddel is.
- **Hoog-risico toepassingen:** AI-systemen die voor bepaalde toepassingen in 'hoog-risico toepassingsgebieden' worden ontwikkeld en ingezet. Dit zijn acht toepassingsgebieden die variëren van AI voor rechtshandhaving tot AI in het onderwijs. Binnen die acht gebieden zijn ongeveer 30 verschillende specifieke toepassingen opgenomen die mogelijk hoge risico's met zich meebrengen, zoals AI-systemen om te helpen bij het uitsturen van nooddiensten.

De productgroepen en toepassingsgebieden waarin AI-systemen als hoog-risico worden aangemerkt, staan in onderstaande figuren.

De verplichtingen voor deze categorie staan beschreven onder 4.2. Hoog-risico AI op pagina 11.

Hoog-risico AI als (veiligheidsonderdeel van) bestaande producten

Dit zijn producten die al gereguleerd zijn in de EU. Een product wordt als risicovol gezien wanneer volgens de bestaande productregelgeving goedkeuring van een derde partij nodig is voordat het in de handel mag worden gebracht (conformiteitsbeoordeling). Als AI een veiligheidsonderdeel van het risicoproduct is of als het risicoproduct zélf een AI-systeem is, dan wordt het als hoog-risico AI gezien. Dit geldt voor producten die onder de volgende productwetgeving vallen:

- **Machines** (Richtlijn 2006/42/EG)
- **Speelgoed** (Richtlijn 2009/48/EG)
- **Pleziervaart** (Richtlijn 2013/53/EU)
- **Liften** (Richtlijn 2014/33/EU)
- **Apparaten en beveiligingssystemen voor gebruik op plaatsen met ontploffingsgevaar** (Richtlijn 2014/34/EU)
- **Radioapparatuur** (Richtlijn 2014/53/EU)
- **Drukapparatuur** (Richtlijn 2014/68/EU)
- **Kabelbaaninstallaties** (Verordening (EU) 2016/424)
- **Persoonlijke beschermingsmiddelen** (Verordening (EU) 2016/425)
- **Gasverbrandingstoestellen** (Verordening (EU) 2016/425)
- **Medische hulpmiddelen** (Verordening (EU) 2017/745)
- **Medische hulpmiddelen voor in-vitrodiagnostiek** (Verordening (EU) 2017/746)

Daarnaast wordt in de AI-verordening nog een lijst producten genoemd die ook worden gezien als hoog-risico AI, maar waar **vanuit de AI-verordening geen directe eisen voor gaan gelden**. Wél worden de eisen uit de AI-verordening op een later moment gebruikt om invulling te geven aan de specifieke productwetgeving die voor deze producten geldt. Wanneer dit gebeurt, is nog niet bekend en zal per product verschillen. Het gaat om de producten die onder de volgende productwetgeving vallen:

- **(Beveiliging van) burgerluchtvaart** (Verordening (EG) 300/2008 en Verordening (EU) 2018/1139)
- **Twee- of driewielige voertuigen en vierwielers** (Verordening (EU) 168/2013)
- **Landbouw- en bosbouwvoertuigen** (Verordening (EU) 167/2013)
- **Uitrusting van zeeschepen** (Richtlijn 2014/90/EU)
- **Interoperabiliteit van het spoorwegsysteem in de EU** (Richtlijn (EU) 2016/797)
- **Motorvoertuigen en aanhangwagens** (Verordening (EU) 2018/858 en Verordening (EU) 2019/2144)

Hoog-risico toepassingsgebieden

Een AI-systeem valt onder een van de toepassingsgebieden als de aanbieder het gebruik van het AI-systeem voor een toepassing binnen één van deze gebieden heeft bedoeld. De aanbieder moet het doel expliciet opnemen in de documenten van het AI-systeem, waaronder de gebruiksinstructies, reclamematerialen en verdere technische documentatie.

Er zijn acht hoog-risico toepassingsgebieden. Dit betekent niet dat alle AI-systemen die binnen dat, vaak abstract omschreven, toepassingsgebied vallen hoog-risico zijn. Per gebied worden een aantal specifieke toepassingen genoemd.

Tip: Ga eerst na of uw AI-systeem onder één van de acht gebieden valt, en kijk vervolgens of uw AI-systeem één van de beschreven AI-systemen in die categorie is. Alleen in dat geval heeft u te maken met een hoog-risico AI-systeem dat aan eisen moet gaan voldoen.

1. **Biometrie**
   - Systemen voor biometrische identificatie op afstand, tenzij het systeem *alleen* voor verificatie gebruikt wordt.
   - Systemen voor biometrische categorisering op basis van gevoelige eigenschappen of kenmerken.
   - Systemen voor emotieherkenning.
2. **Kritieke infrastructuur**
   - Systemen als veiligheidscomponent voor beheer en exploitatie van kritieke digitale infrastructuur, voor wegverkeer of voor levering van water, gas, verwarming en elektriciteit.
3. **Onderwijs en beroepsopleiding**
   - Systemen voor toelating tot of toewijzing van onderwijs.
   - Systemen voor het evalueren van leerresultaten.
   - Systemen voor het beoordelen van het onderwijsniveau.
   - Systemen voor het monitoren van leerlingen tijdens examens.
4. **Werkgelegenheid, personeelsbeheer en toegang tot zelfstandige arbeid**
   - Systemen voor het werven of selecteren van kandidaten.
   - Systemen voor het maken van besluiten die gevolgen hebben voor het arbeidscontract en de arbeidsvoorwaarden, voor het verdelen van taken en voor het monitoren en evalueren van werknemers.
5. **Essentiële particuliere en publieke diensten en uitkeringen**
   - Systemen voor het beoordelen van (de mate van) toegang tot essentiële overheidsuitkeringen en -diensten.
   - Systemen voor het beoordelen van de kredietwaardigheid of -score van personen, tenzij dit gebeurt om financiële fraude op te sporen.
   - Systemen voor de risicobeoordeling en prijsbepaling voor levens- en ziektekostenverzekeringen.
   - Systemen voor het prioriteren van de inzet van hulpdiensten of voor de triage van patiënten in de zorg.
6. **Rechtshandhaving**
   - Systemen voor rechtshandhaving om te bepalen hoe groot de kans is dat iemand slachtoffer van een misdrijf wordt.
   - Systemen voor rechtshandhaving die worden ingezet als leugendetector.
   - Systemen voor rechtshandhaving om betrouwbaarheid van bewijsmateriaal te beoordelen.
   - Systemen voor rechtshandhaving om te beoordelen of te voorspellen hoe groot de kans is dat iemand een strafbaar feit zal plegen of om eerder crimineel gedrag van (groepen) personen te beoordelen.
   - Systemen voor rechtshandhaving voor het profileren van personen tijdens opsporing of vervolging.
7. **Migratie-, asiel- en grensbeheer**
   - Systemen voor overheidsinstanties om te gebruiken als leugendetector.
   - Systemen voor overheidsinstanties om risico's voor veiligheid, op illegale migratie of een gezondheidsrisico bij binnenkomst van een land te beoordelen.
   - Systemen om overheidsinstanties bij te staan bij het behandelen van asiel, visa of verblijfsvergunning, inclusief aanverwante klachten.
   - Systemen voor overheidsinstanties voor het opsporen, herkennen of identificeren van personen, met uitzondering van de verificatie van reisdocumenten.
8. **Rechtspraak en democratische processen**
   - Systemen om de rechtspraak en geschillenbeslechting te ondersteunen bij het onderzoeken en uitleggen van feiten of om de wet uit te leggen of toe te passen op een set aan feiten.
   - Systemen voor het beïnvloeden van verkiezingen of referenda of het stemgedrag van personen, met uitzondering van de ondersteuning van politieke campagnes uit administratief of logistiek oogpunt.

Uitzonderingen op de hoog-risico toepassingsgebieden
Er zijn een aantal specifieke **uitzonderingen** waarbij AI-systemen die onder één van de toepassingsgebieden vallen tóch niet als hoog-risico AI worden gezien. Dat is het geval wanneer het AI-systeem **geen wezenlijke invloed heeft op een besluit**, omdat het systeem bedoeld is om:

- Een **beperkte procedurele taak** uit te voeren;
- Het resultaat van een **eerder voltooide menselijke activiteit te verbeteren**;
- Om te **controleren of menselijke besluiten afwijken** van een eerder patroon zonder het besluit te vervangen of te beïnvloeden;
- Een **voorbereidende taak** uit te voeren voor een beoordeling die relevant is voor een van de hoog-risico toepassingsgebieden.

Als u heeft vastgesteld dat uw AI-systeem onder één van de uitzonderingen valt, moet u dit vastleggen en het AI-systeem registreren in de EU-databank voor hoog-risico AI-systemen. De Europese Commissie stelt op een later moment nog een lijst met voorbeelden op om te verduidelijken wat wel en niet onder de uitzonderingen valt.

1.3. AI modellen en AI-systemen voor algemene doeleinden

Een **AI-model voor algemene doeleinden**, ook wel bekend als 'General Purpose AI', kan een breed scala aan verschillende taken op een goede manier uitvoeren en kan daardoor worden geïntegreerd in allerlei verschillende AI-systemen. Vaak worden deze modellen getraind op grote hoeveelheden data en met behulp van *self-supervision* technieken.

De brede inzetbaarheid van deze modellen, via specifieke AI-systemen, zorgt ervoor dat ze voor allerlei soorten toepassingen worden gebruikt. Dit kunnen ook hoog-risico toepassingen zijn. Vanwege de potentieel grote impact van deze modellen, moeten ze vanaf augustus 2025 aan verschillende eisen voldoen.

Als een AI-systeem gebaseerd is op een AI-model voor algemene doeleinden en dat ook zelf daadwerkelijk verschillende doeleinden kan dienen, dan is het een **AI systeem voor algemene doeleinden**.

De verplichtingen voor deze categorie staan beschreven onder 4.3. AI-modellen en systemen voor algemene doeleinden op pagina 16.

1.4. Generatieve AI en chatbots

Aan sommige AI-systemen worden transparantieverplichtingen gesteld. Dit zijn systemen waarmee mensen vaak direct in aanraking komen. Daarom moet het voor mensen duidelijk zijn dat er sprake is van interactie met AI of dat de content is gemanipuleerd of gegenereerd.

- Systemen die gebruikt worden om audio, beeld, video of tekst te genereren (**generatieve AI**);
- Systemen die gemaakt zijn voor interactie (**chatbots**).

De verplichtingen voor deze categorie staan beschreven onder 4.4. Generatieve AI en Chatbots op pagina 17.

1.5. Overige AI

Zie 4.5. Overige AI op pagina 17 voor meer informatie over AI-systemen die niet onder één van de hierboven beschreven risicocategorieën vallen.

Stap 2. Is ons systeem 'AI' volgens de AI-verordening?

De AI-verordening stelt eisen aan AI-systemen. Er bestaan verschillende beelden over wat 'AI' wel of niet is. In de AI-verordening staat de volgende definitie die bedoeld is om af te bakenen wat AI is als product op de markt:

"Een AI-systeem is een op een machine gebaseerd systeem dat is ontworpen om ***met verschillende niveaus van autonomie*** *te werken en dat na de uitrol* ***aanpassingsvermogen*** *kan vertonen, en dat, voor* ***expliciete of impliciete doelstellingen****, uit de ontvangen input afleidt hoe* ***output*** *te genereren zoals voorspellingen, inhoud, aanbevelingen of beslissingen."*

Wat valt hier onder andere onder?

- Systemen die gebruik maken van machinaal leren (***machine learning***) waarbij op basis van data wordt geleerd hoe bepaalde doelstellingen kunnen worden bereikt;
- Systemen die gebruik maken van op logica en kennis gebaseerde benaderingen (***knowledge and logic-based approaches***) die leren, redeneren of modelleren mogelijk maken.

Wat valt hier **niet** onder?

- Systemen die gebaseerd zijn op regels die uitsluitend door mensen zijn vastgesteld om automatisch handelingen uit te voeren.

Als uw systeem niet gezien wordt als 'AI' onder de AI-verordening maar wel onder één van de risicocategorieën valt, is het belangrijk om binnen uw organisatie het gesprek te voeren in hoeverre het systeem niet toch risico's vormt en of deze risico's te mitigeren zijn door aan (bepaalde) eisen uit de AI-verordening te voldoen. Daarnaast moeten de systemen die buiten de AI-verordening vallen mogelijk wel aan eisen uit andere regelgeving voldoen.

Stap 3. Zijn wij de aanbieder of gebruiksverantwoordelijke van het AI-systeem?

Nadat u heeft bepaald onder welke risicocategorie uw AI-systeem valt en of uw AI-systeem ook onder de AI-verordening valt, moet u bepalen of u een aanbieder of een gebruiksverantwoordelijke bent.

- **Aanbieder (provider)**: een persoon of organisatie die een AI-systeem of model ontwikkelt of laat ontwikkelen en dat systeem in de handel brengt of zelf in gebruik neemt.
- **Gebruiksverantwoordelijke (deployer)**: een persoon of organisatie die een AI-systeem onder eigen verantwoordelijkheid gebruikt. Niet-professioneel gebruik valt hier niet onder.

In de beschrijving van de eisen in stap 4 wordt per risicocategorie beschreven welke verplichtingen gelden voor aanbieders en gebruiksverantwoordelijken. Zij moeten elk aan andere verplichtingen voldoen, waarbij de zwaarste verplichtingen voor aanbieders gelden.

Let op: Als gebruiksverantwoordelijke kunt u in sommige gevallen ook aanbieder worden van een hoog-risico AI-systeem en moet u voldoen aan de hoog-risico verplichtingen voor aanbieders. Dit wordt verder toegelicht onder stappen 4.2. Hoog-risico AI op pagina 11 en 4.3. AI-modellen en systemen voor algemene doeleinden op pagina 16.

Let op: er zijn ook andere rollen onder de AI-verordening, bijvoorbeeld gemachtigden, importeurs en distributeurs. Op verplichtingen voor deze actoren gaat deze gids niet in.

Stap 4. Aan welke verplichtingen moeten wij ons houden?

4.1. Verboden AI

Deze AI-systemen brengen een onaanvaardbaar risico met zich mee en zijn daarom vanaf 2 februari 2025 verboden. Dit betekent dat deze systemen niet in de handel gebracht of gebruikt mogen worden. Deze verboden gelden zowel voor **aanbieders** als **gebruiksverantwoordelijken**.

Voor het verbod op biometrische identificatie op afstand in real time in openbare ruimten met het oog op de rechtshandhaving bestaan scherp afgebakende uitzonderingen en de inzet ervan dient te worden voorzien van een nationale wettelijke grondslag. Ook gelden er aanvullende waarborgen rond de inzet van deze systemen.

4.2. Hoog-risico AI

De zwaarste verplichtingen uit de AI-verordening gaan gelden voor hoog-risico AI-systemen. **Aanbieders** moeten aan (verschillende) verplichtingen voldoen:

- Systeem voor risicomanagement;
- Data en datagovernance;
- Technische documentatie;
- Logging;
- Transparantie en informatie;
- Menselijk toezicht;
- Nauwkeurigheid, robuustheid en cyberbeveiliging;
- Systeem voor kwaliteitsbeheer;
- Monitoring.

Als u als aanbieder aan al deze verplichtingen voldoet of denkt te voldoen, zult u een **conformiteitsbeoordeling** moeten uitvoeren. In sommige gevallen mag u dit zelf doen en in sommige gevallen moet een derde partij dit voor u doen. Een latere versie van deze gids zal ingaan op wanneer u welke procedure moet uitvoeren.

Ook **gebruiksverantwoordelijken** moeten aan verschillende verplichtingen voldoen, waarbij er extra verplichtingen gelden voor overheidsorganisaties die AI-systemen gebruiken.

In het figuur hieronder wordt elke verplichting toegelicht. Deze verplichtingen worden komende jaren verder uitgewerkt in Europese normen, waaraan iedereen kan deelnemen via de normalisatieorganisaties in de Europese lidstaten. In Nederland is dit het NEN.[3]

Let op: In twee gevallen kunt u als gebruiksverantwoordelijke van een hoog-risico AI-systeem zélf de aanbieder van dat systeem worden:

- Wanneer u als gebruiksverantwoordelijke uw eigen naam of merk op het hoog-risico systeem zet;
- Wanneer u als gebruiksverantwoordelijke een aanzienlijke wijziging aan het hoog-risico AI-systeem aanbrengt die niet door de aanbieder was voorzien en waardoor het systeem niet meer aan de eisen voldoet of waardoor het door de aanbieder beoogde doel van het systeem verandert.

[3] https://www.nen.nl/ict/digitale-ehtiek-en-veiligheid/ai

Regels voor aanbieders van hoog-risico AI-systemen

1. Systeem voor risicomanagement
Er moeten verschillende stappen worden ondernomen voor dit systeem:

- Het vaststellen en analyseren van voorzienbare risico's voor onder andere gezondheid, veiligheid of grondrechten.
- Het nemen van geschikte risicobeheersmaatregelen die ervoor zorgen dat risico's die na de maatregelen overblijven *aanvaardbaar* zijn.

Hierbij moet rekening gehouden met de volgende punten:
- Het vaststellen en aanpakken van risico's moet eenmalig plaatsvinden voordat het AI-systeem in de handel gebracht of gebruikt wordt en daarna continu tijdens het gebruik van het AI-systeem.
- Er moet rekening gehouden worden met voorzienbaar misbruik van het systeem.
- Er moet rekening gehouden worden met de context van het gebruik, waaronder de kennis en ervaring van de gebruiksverantwoordelijke met dergelijke AI-systemen of het feit dat kinderen of kwetsbare groepen gevolgen van het AI-systeem ondervinden. Het kan bijvoorbeeld nodig zijn om een opleiding aan te bieden aan degenen die met het AI-systeem gaan werken.
- De risicobeheersmaatregelen moeten getest worden om te controleren of ze ook daadwerkelijk werken. Dit moet gebeuren aan de hand van benchmarks die passen bij het doel waarvoor het AI-systeem wordt ingezet.
- Als onder bestaande productwetgeving ook een risicobeheersysteem moet worden opgesteld, dan mag het samengevoegd worden tot één risicobeheersysteem.

2. Data en datagovernance
Er worden verschillende eisen gesteld aan de datasets die voor het trainen, valideren en testen van hoog-risico AI-systemen worden gebruikt.

- Databeheer die past bij het doel van het AI-systeem, waaronder:
  - Het registreren van de processen, waaronder die van dataverzameling en -verwerking;
  - Het vastleggen van aannames over de datasets;
  - Een beoordeling van de beschikbaarheid, hoeveelheid en geschiktheid van de datasets, waaronder mogelijke vertekeningen die gevolgen voor mensen kunnen hebben;
  - Maatregelen om vertekeningen op te sporen, te voorkomen en te beperken;
  - Het aanpakken van tekortkomingen in de datasets die het voldoen aan de AI-verordening in de weg kunnen zitten (denk bijvoorbeeld aan het beperken van risico's onder het risicomanagementsysteem).
- Datasets moeten voor het doel waarvoor ze gebruikt worden voldoende representatief en zoveel mogelijk foutenvrij zijn. Hierbij moet ook rekening gehouden worden met de context waarin het AI-systeem gaat worden gebruikt; bijvoorbeeld de geografische of sociale context.
- Onder een aantal strenge voorwaarden mogen bijzondere categorieën persoonsgegevens (een begrip uit de Algemene Verordening Gegevensbescherming) verwerkt worden om vertekeningen in de datasets tegen te gaan.

3. Technische documentatie
In de technische documentatie moet worden aangetoond dat het hoog-risico AI-systeem voldoet aan de eisen uit de AI-verordening. In de technische documentatie moet onder andere worden opgenomen:

- Een algemene beschrijving van het AI-systeem, waaronder het beoogde doel van het systeem, de naam van de aanbieder en de instructies voor gebruik;
- Een gedetailleerde beschrijving van de elementen van het AI-systeem en van het proces voor de ontwikkeling ervan, waaronder de stappen van ontwikkeling, de ontwerpkeuzen, de verwachte output van het systeem, het risicobeheersysteem en de gebruikte datasets.

- Gedetailleerde informatie over de monitoring, werking en controle van het AI-systeem, waaronder de mate van nauwkeurigheid op individueel en algemeen niveau, risico's, het systeem voor evaluatie tijdens gebruik en maatregelen voor monitoring en menselijk toezicht.
- Een overzicht van de standaarden die worden toegepast.
- De EU-conformiteitsverklaring (het CE-keurmerk).

MKB-bedrijven kunnen de technische documentatie op een versimpelde manier vastleggen. De Europese Commissie verstrekt hier op een later moment een formulier voor.

4. Logging
Er moeten automatisch logs worden bijgehouden tijdens de levensduur van het AI-systeem om hiermee risico's tijdig op te kunnen sporen en de werking van het systeem te kunnen monitoren. De logs moeten voor ten minste zes maanden bewaard worden. Er moet in ieder geval het volgende gelogd worden:
- De duur van elk gebruik van het AI-systeem;
- De inputdata en de controle daarvan door het AI-systeem (en de referentiedatabank);
- De identificatie van de personen die betrokken zijn bij de verificatie van de resultaten.

5. Transparantie en informatie
De aanbieder van het AI-systeem weet hoe het systeem werkt en hoe hiermee om gegaan moet worden. Zij moet er daarom voor zorgen dat het AI-systeem zodanig transparant is dat gebruiksverantwoordelijken begrijpen hoe zij de output ervan op een goede manier kunnen gebruiken.

Hiervoor moeten **instructies voor gebruik** opgesteld worden, die onder andere de volgende punten bevat:

- Contactgegevens;
- Het doel, de kenmerken, de capaciteiten en de beperkingen van de prestaties van het AI-systeem;
- De maatregelen voor menselijk toezicht.

6. Menselijk toezicht
Hoog-risico AI-systemen moeten door de aanbieder zo worden ontworpen dat er tijdens het gebruik menselijk toezicht op kan worden uitgeoefend, om hiermee risico's voor personen te beperken. Het menselijk toezicht is contextafhankelijk – des te groter de risico's des te sterker het toezicht moet zijn. De maatregelen voor het toezicht kunnen technisch van aard zijn (bijvoorbeeld een duidelijke mens-machine interface), of maatregelen die de gebruiksverantwoordelijken moeten uitvoeren (bijvoorbeeld een verplichte cursus voor hun personeel).

Uiteindelijk moeten de maatregelen ervoor zorgen dat de personen die het AI-systeem gaan gebruiken, het volgende kunnen:

- De capaciteiten van het systeem begrijpen en de werking monitoren;
- Bewust zijn van 'automation bias';
- De output juist interpreteren en indien nodig negeren of vervangen;
- Het systeem stopzetten.

7. Nauwkeurigheid, robuustheid en cyberbeveiliging
Hoog-risico AI-systemen moeten een passend niveau van nauwkeurigheid, robuustheid en cyberbeveiliging bieden. Hiervoor worden, onder andere door de Europese Commissie, benchmarks en meetmethoden ontwikkeld.

Hiervoor worden in ieder geval de volgende maatregelen genoemd:

- Technische en organisatorische maatregelen om fouten te voorkomen die ontstaan bij het gebruik van het AI-systeem door personen;
- Oplossingen voor robuustheid, zoals back-ups of veiligheidsmaatregelen bij defecten;
- Negatieve beïnvloeding van het systeem wegnemen of beperken door feedback loops te beperken;
- Cyberbeveiliging die ongewenste toegang door derden voorkomt, door aanvallen te traceren, erop te reageren en op te lossen. Het gaat dan om aanvallen die gericht zijn op datavervuiling, modelvervuiling, aanpassen van input of het verkrijgen van vertrouwelijke gegevens.

8. Systeem voor kwaliteitsbeheer
Het systeem van kwaliteitsbeheer moet ervoor zorgen dat de eisen uit de AI-verordening worden nageleefd. Hoe uitgebreid dit systeem van kwaliteitsbeheer moet zijn, is afhankelijk van de grootte van de organisatie. Onder andere door het vastleggen van :

- Een strategie voor naleving;
- Technieken, procedures en maatregelen voor het ontwerp, de ontwikkeling en kwaliteitscontrole van het AI-systeem;
- Het wel of niet gebruiken van standaardisatie;
- Systemen en procedures voor databeheer, risicobeheer, monitoring, incidentmelding en documentatie.

9. Monitoring
Zodra een AI-systeem in de handel gebracht is of in gebruik is, moeten aanbieders het systeem monitoren op basis van gebruiksdata om hiermee te controleren of het systeem aan de eisen uit de AI-verordening blijft voldoen. Hiervoor moeten aanbieders een plan voor monitoring opstellen.

Als de aanbieder van het hoog-risico AI-systeem erachter komt dat het systeem niet meer in overeenstemming met de AI-verordening werkt, moeten er meteen corrigerende maatregelen genomen worden om dit te herstellen. Dit kan zelfs het terugroepen van het systeem zijn indien dat nodig is. Ook moet de aanbieder met gebruiksverantwoordelijken samenwerken en de toezichthouders informeren hierover.

Ernstige incidenten met het AI-systeem moeten bij de toezichthouder(s) gemeld worden.

Overige eisen
- De registratie van het hoog-risico AI-systeem in de EU-databank.
- De contactgegevens van de aanbieder moeten bij het AI-systeem vermeld worden;
- De technische documentatie, documentatie rondom het kwaliteitsbeheer en documentatie rondom de conformiteitsbeoordeling moeten voor 10 jaar worden bewaard.

Regels voor gebruiksverantwoordelijken van hoog-risico AI-systemen
Niet alleen aanbieders, maar ook de gebruiksverantwoordelijken van hoog-risico AI-systemen moeten aan eisen gaan voldoen. Zij zijn immers degenen die controle hebben over hoe het AI-systeem in de praktijk gebruikt wordt en hebben daarmee een grote invloed op de risico's die kunnen ontstaan.

Gebruiksverantwoordelijken moeten:

- Technische en organisatorische maatregelen nemen om ervoor te zorgen dat het hoog-risico AI-systeem volgens de gebruiksinstructies wordt gebruikt;
- Ervoor zorgen dat de personen die toezicht houden op het systeem over de juiste kennis, kunde en autoriteit beschikken;
- Ervoor zorgen dat, voor zover dat kan, de inputdata voldoende relevant en representatief is;
- De werking van het AI-systeem monitoren op basis van de gebruiksinstructies;
- Als een gebruiksverantwoordelijke aanneemt dat het systeem niet meer aan de eisen uit de AI-verordening voldoet, dient deze de aanbieder op de hoogte te stellen en gebruik van het systeem te onderbreken;
- De aanbieder en toezichthouder(s) op de hoogte stellen van mogelijke risico's en ernstige incidenten die hebben plaatsgevonden;
- De logs waarover zij controle hebben voor ten minste zes maanden bewaren;
- De werknemersvertegenwoordiging informeren indien het AI-systeem op de werkvloer wordt ingezet;
- Als er beslissingen over mensen gemaakt worden met behulp van het hoog-risico AI-systeem, deze mensen hierover informeren.
- Als er gebruik wordt gemaakt van AI voor emotieherkenning of biometrische categorisatie, moeten de mensen op wie dit wordt toegepast hierover geïnformeerd worden.

Specifieke verplichtingen voor overheidsorganisaties als gebruiksverantwoordelijken
Overheidsorganisaties moeten in aanvulling op bovenstaande verplichtingen aan een aantal extra verplichtingen voldoen:

- Het registreren van het gebruik van een hoog-risico systeem in de EU-databank;
- Beoordelen wat de gevolgen voor de grondrechten kunnen zijn bij het gebruik van het hoog-risico AI-systeem, met oog voor de specifieke context waarbinnen het gebruik plaats vindt (een **'fundamental rights impact assessment'**). Zij kijken daarbij bijvoorbeeld naar de duur van het gebruik, de processen waarbinnen het systeem gebruikt wordt, en de impact die het gebruik op de grondrechten van natuurlijke personen en groepen kan hebben. Na identificatie van de risico's moeten gebruiksverantwoordelijkheden maatregelen treffen voor menselijk toezicht en het omgaan met de eventuele risico's. Ook moet er een melding bij de markttoezichthouder gedaan worden.
  Let op: Deze verplichting geldt ook voor particulieren die openbare diensten verlenen, het gebruik van AI-systemen voor het beoordelen van de kredietwaardigheid van mensen en AI-systemen voor risicobeoordelingen voor levens -en ziektekostenverzekeringen.

4.3. AI-modellen en systemen voor algemene doeleinden

Verplichtingen voor aanbieders van AI-modellen voor algemene doeleinden
AI-modellen voor algemene doeleinden kunnen in allerlei verschillende AI-systemen geïntegreerd worden. Hiervoor is het wel van belang dat de aanbieders van deze AI-systemen weten wat het AI-model wel en niet kan. Ook worden er specifieke eisen gesteld aan het trainen van deze modellen, omdat hier vaak grote datasets voor gebruikt worden. De aanbieders van deze modellen moeten:

- Technische documentatie opstellen over onder andere het trainings- en testproces en de resultaten en evaluatie daarvan;
- Informatie en documentatie opstellen en updaten voor aanbieders van AI-systemen die het model in hun AI-systeem willen integreren. De informatie moet inzicht geven in de capaciteiten en beperkingen van het AI-model en moet de aanbieder van het AI-systeem in staat stellen om zelf aan de verplichtingen uit de AI-verordening te kunnen voldoen.
- Beleid opstellen om ervoor te zorgen dat ze het model trainen zonder inbreuk te maken op de auteursrechten van mensen en organisaties;
- Een voldoende gedetailleerde samenvatting opstellen en openbaar maken over de content die gebruikt is om het AI-model te trainen.

Aanbieders van open-source modellen hoeven niet aan deze verplichtingen te voldoen. De AI-verordening bepaalt wanneer een model 'echt' open-source is.

Verplichtingen voor aanbieders van AI-modellen voor algemene doeleinden met systeemrisico's
In sommige gevallen kunnen AI-modellen voor algemene doeleinden voor systeemrisico's zorgen. Dit is het geval als het model capaciteiten met een grote impact heeft. Er wordt in ieder geval vanuit gegaan dat dit zo is, wanneer er ten minste 10^{25} 'floating-point operations' (FLOPs) zijn gebruikt om het model te trainen. Daarnaast kan de Europese Commissie op basis van bepaalde criteria vaststellen dat het model op een andere manier een vergelijkbaar grote impact heeft. Deze modellen moeten:

- Aan de verplichtingen voor AI-modellen voor algemene doeleinden voldoen;
- Modelevaluaties uitvoeren om systeemrisico's in kaart te brengen;
- Systeemrisico's beperken;
- Informatie over ernstige incidenten bijhouden en rapporteren aan het AI-Office;
- Zorgen voor passende cyberbeveiliging.

Let op: deze verplichtingen gelden alleen voor de grootste AI-modellen.

Aanbieders van deze modellen met systeemrisico's kunnen zich niet op een uitzondering voor open-source beroepen.

Welke rechten heeft u als u een AI-model voor algemene doeleinden integreert in uw (hoog-risico) AI systeem? Zoals hierboven ook aangegeven, moet u in ieder geval informatie en documentatie ontvangen zodat u zelf een inschatting kunt maken op welke manier u het model in uw AI-systeem kunt gebruiken voor het door u gekozen doel. Als u het model in een hoog-risico AI-systeem verwerkt, moet u zich vervolgens als aanbieder nog steeds aan de verplichtingen uit de AI-verordening houden.

Hoe moet u omgaan met AI-systemen voor algemene doeleinden? Zoals aangegeven onder 1.3. AI modellen en AI-systemen voor algemene doeleinden op pagina 8, bestaan er ook AI-systemen die

verschillende doeleinden kunnen dienen. Denk bijvoorbeeld aan de bekende AI-chatbots. **Let op:** Indien u deze systemen voor hoog-risico doeleinden inzet, wordt u volgens de AI-verordening zelf de aanbieder van een hoog-risico AI-systeem. U zal dan aan de bijbehorende verplichtingen moet voldoen. Het is in deze situatie erg lastig om aan de verplichtingen voor een hoog-risico AI-systeem te voldoen, waardoor u het risico loopt om een boete te ontvangen.

4.4. Generatieve AI en Chatbots

Om ervoor te zorgen dat mensen weten of zij met een AI-systeem praten of content zien die door AI is gegenereerd, worden er transparantieverplichtingen gesteld aan generatieve AI en chatbots.

Regels voor aanbieders van chatbots
Aanbieders van systemen die gemaakt zijn voor directe interactie met mensen, moeten ervoor zorgen dat deze mensen erover worden geïnformeerd dat ze met een AI-systeem te maken hebben.

Regels voor aanbieders van generatieve AI
Aanbieders van systemen die audio, beeld, video of tekst genereren moeten ervoor zorgen dat de output in een machineleesbaar format gemarkeerd wordt zodat de output herkend kan worden als kunstmatig gegenereerd of gemanipuleerd.

Regels voor gebruiksverantwoordelijken van generatieve AI
Gebruiksverantwoordelijken van systemen die audio, beeld, video genereren moeten ervoor zorgen dat duidelijk is dat de inhoud kunstmatig gegenereerd of gemanipuleerd is. Dit kan bijvoorbeeld door een watermerk. Voor creatief, satirisch, fictief of analoog werk mag dit gedaan worden op een manier die het werk niet verpest.

Voor kunstmatig gegenereerde tekst geldt een speciaal regime. Alleen voor tekst die gebruikt wordt om ‘het publiek te informeren over aangelegenheden van algemeen belang’ moet er duidelijk gemaakt worden dat het kunstmatig is gegenereerd of bewerkt. Wanneer er sprake is van redactionele controle en verantwoordelijkheid, hoeft dit niet gedaan te worden.

Regels voor gebruiksverantwoordelijken van emotieherkenningssystemen of systemen voor biometrische categorisering
De gebruiksverantwoordelijken van deze AI-systemen moeten de mensen die aan het systeem blootgesteld worden informeren over hoe het systeem werkt.

4.5. Overige AI

AI-systemen die buiten de hierboven genoemde categorieën vallen, hoeven niet aan eisen onder de AI-verordening te voldoen.

Maar let op: als u als **gebruiksverantwoordelijke** het ‘overige categorie’ AI-systeem inzet voor een hoog-risico toepassing zoals genoemd in de AI-verordening (zie 1.2. Hoog-risico AI-systemen op pagina 5), dan wordt het automatisch een hoog-risico AI-systeem en moet u als **aanbieder** van het systeem voldoen aan de eisen uit de AI-verordening.

Deze brochure is een uitgave van:

Rijksoverheid
Postbus 00000 | 2500 AA Den Haag
T 0800 646 39 51 (ma t/m vrij 9.00 – 21.00 uur)

Oktober 2024 | Publicatie-nr. 104969