Knowledge Centre
Data & Society

Deep Dive: AI Act

Thomas Gils
KU Leuven Centre for IT and IP Law / Kenniscentrum
Data & Maatschappij
7 juni 2022

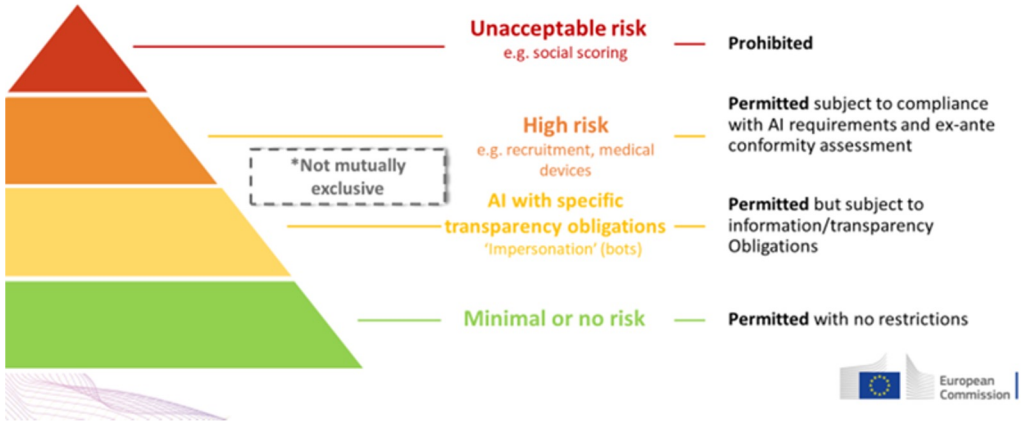
Intro AI Act

Kenniscentrum Data & Maatschappij

04/2018: *AI for Europe* → 02/2020: *Witboek AI* → 04/2021: *Voorstel AI Act*

- Witboek AI: “Ecosysteem van vertrouwen”
- Voorstel AI Act:

A risk-based approach to regulation



Intro AI Act

Kenniscentrum Data & Maatschappij

- Definitie AI: 3 elementen

I.	II.	III.
Software	Ontwikkeld aan de hand van **bepaalde technieken**	Die voor een bepaalde reeks door mensen gedefinieerde doelstellingen **output** kan genereren die van invloed is op de omgeving waarmee wordt geïnterageerd

Annex I AIA:

- ✓ *alle vormen van ML*
- ✓ *logica-gebaseerde benaderingen (incl. bv. expertsystemen)*
- ✓ *statistische benaderingen*

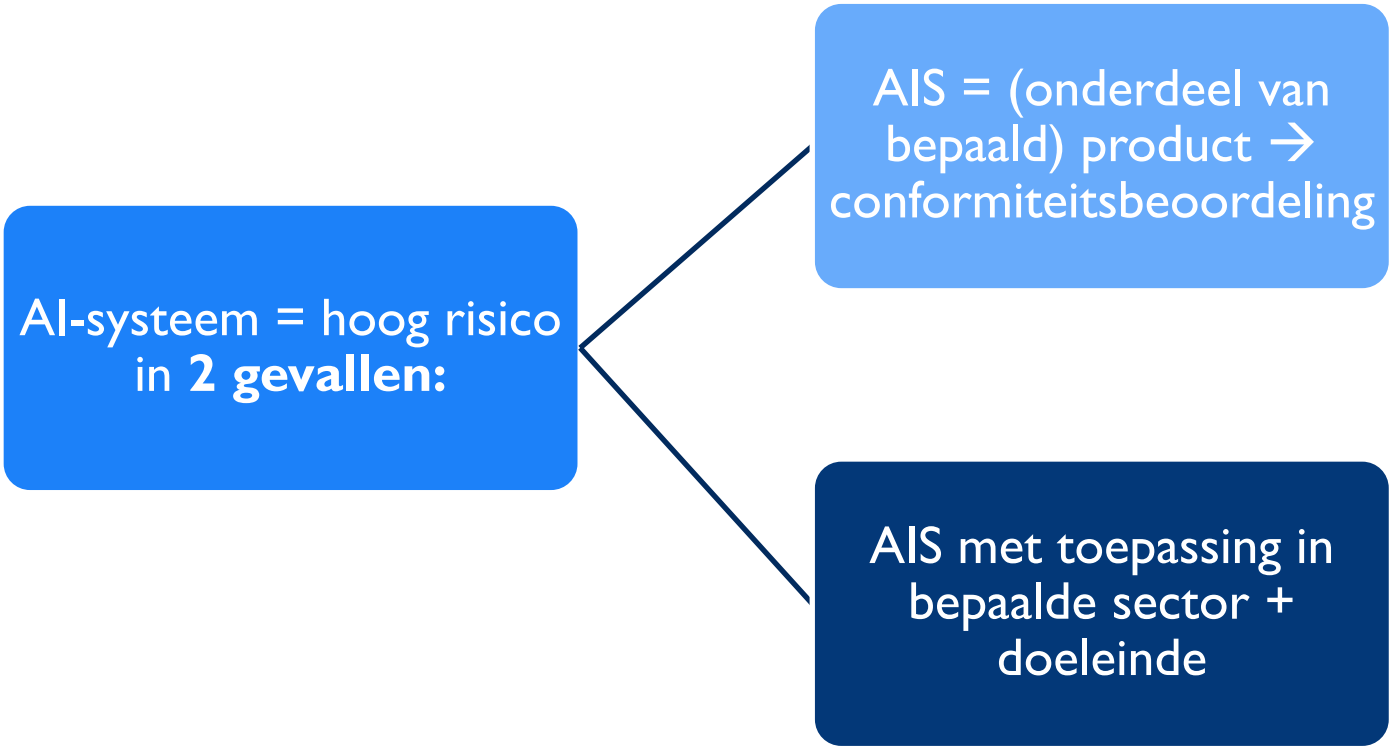
Output: *inhoud, voorspellingen, aanbevelingen of beslissingen,…*

Kenniscentrum
Data & Maatschappij

Hoog Risico AI-systemen?

Hoog Risico AI-systemen

Kenniscentrum Data & Maatschappij



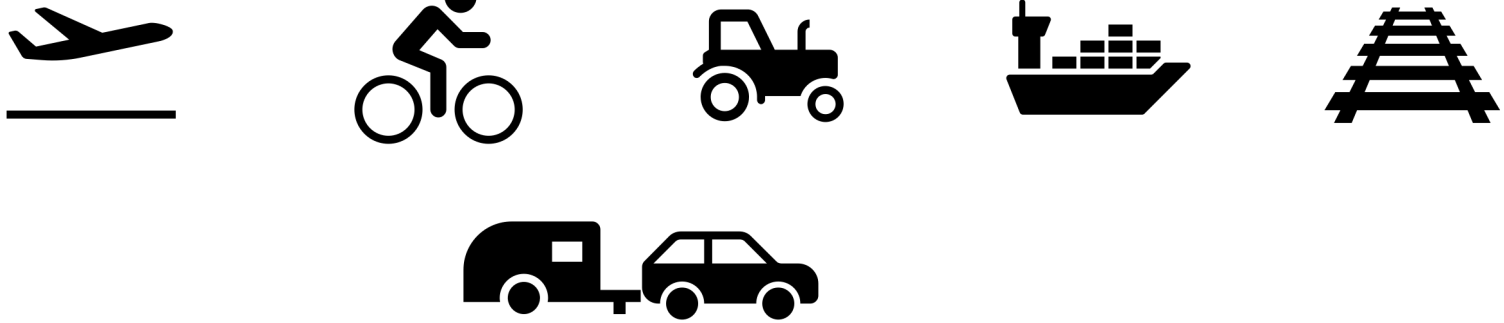
Type 1 Hoog Risico

Kenniscentrum Data & Maatschappij

AI = hoog risico indien:

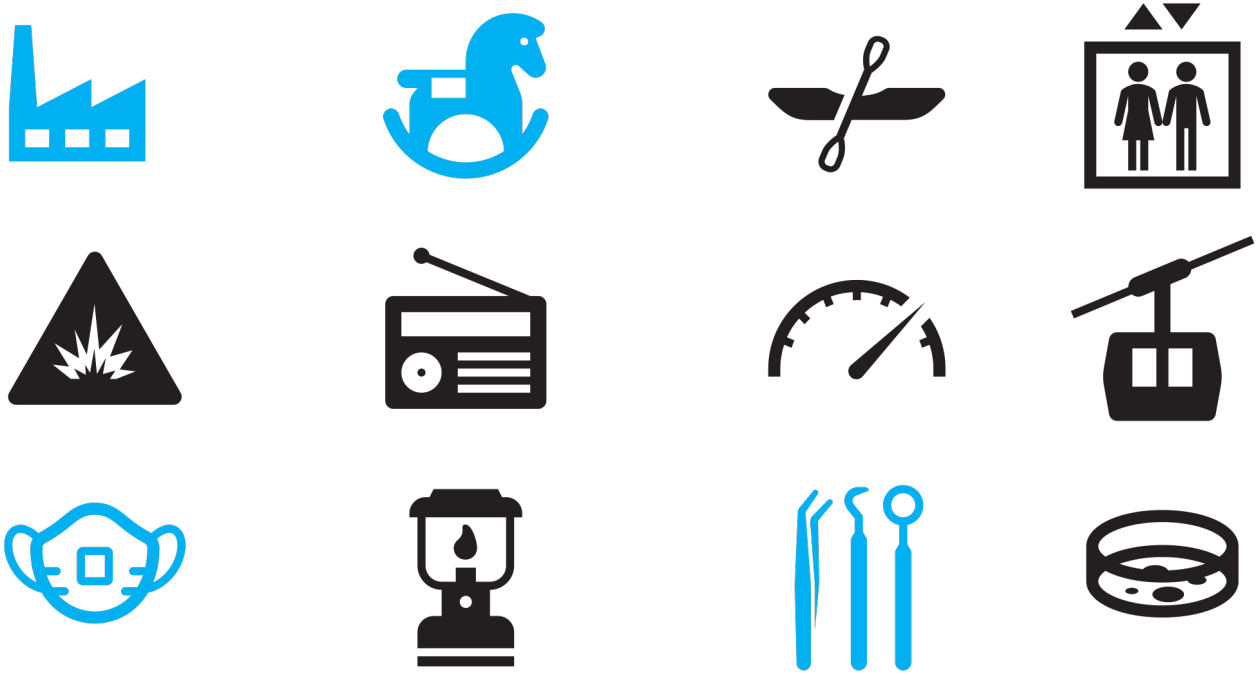
- AIS = **product of VC v/e product** dat onder deel A van de lijst van wetgeving in Annex II valt
- **EN** dit product vereist conformiteitsbeoordeling door 3e partij

Welke producten niet? (= Deel B van Annex II)



Welke producten wel?

Kenniscentrum Data & Maatschappij



Type 2 Hoog Risico

Kenniscentrum Data & Maatschappij

- **Biometrische identificatie en categorisering van natuurlijke personen**
  - Op afstand van natuurlijke personen in real time en achteraf
- **Beheer en exploitatie van kritieke infrastructuur:**
  - VC bij beheer of exploitatie van wegverkeer en de levering van water, gas, verwarming en elektriciteit
- **Onderwijs en beroepsopleiding:**
  - Bepaling van toegang tot of de toewijzing aan onderwijsinstellingen
  - Beoordeling van studenten en deelnemers aan toelatingstoetsen
- **Werkgelegenheid, personeelsbeheer en toegang tot zelfstandige arbeid:**
  - Aanwerving of selectie van kandidaten (incl. screenen of filteren van sollicitaties, de evaluatie van kandidaten)
  - Beslissingen over de bevordering en beëindiging van arbeidsrelaties, voor de toewijzing van taken en voor het toezicht/evaluatie van arbeidsprestaties

Type 2 Hoog Risico (2)

Kenniscentrum Data & Maatschappij

- **Toegang tot en gebruik van essentiële particuliere diensten en openbare diensten en uitkeringen**:
  - Gebruik door of namens overheidsinstanties om te beoordelen of burgers in aanmerking komen voor overheidsuitkeringen en –diensten (incl. intrekken etc.)
  - Beoordeling van kredietwaardigheid van natuurlijke personen
  - Gebruik voor de inzet of voor het vaststellen van prioriteiten bij de inzet van hulpdiensten
- **Rechtshandhaving** *(beperkt tot rechtshandhavingsautoriteiten)*:
  - Gebruik voor het uitvoeren van individuele risicobeoordelingen van natuurlijke personen om te beoordelen hoe groot het risico is dat een natuurlijke persoon (opnieuw) een strafbaar feit pleegt of hoe groot het risico is voor mogelijke slachtoffers van strafbare feiten;
  - Leugendetectoren of soortgelijke hulpmiddelen ter vaststelling van de emotionele toestand van een natuurlijke persoon;
  - Gebruik voor de opsporing van deep fakes,
  - Gebruik voor de beoordeling van de betrouwbaarheid van bewijsmateriaal
  - Gebruik voor het voorspellen van een daadwerkelijk of potentieel strafbaar feit dat (opnieuw) zal worden gepleegd, op basis van de profilering van natuurlijke personen of de beoordeling van persoonlijkheidskenmerken en kenmerken of eerder crimineel gedrag van natuurlijke personen of groepen;
  - Gebruik voor de profilering van natuurlijke personen tijdens de opsporing, het onderzoek of vervolging van strafbare feiten; […]

Type 2 Hoog Risico (3)

Kenniscentrum Data & Maatschappij

- **Migratie, asiel en beheer van grenscontroles** *(beperkt tot door bevoegde overheidsinstanties):*
  - Leugendetectoren of soortgelijke hulpmiddelen voor de vaststelling van de emotionele toestand van een natuurlijke persoon;
  - Gebruik voor de beoordeling van een risico, waaronder een beveiligingsrisico, een risico op illegale immigratie of een gezondheidsrisico, dat een natuurlijke persoon vormt die voornemens is het grondgebied van een lidstaat te betreden of dat heeft gedaan;
  - Gebruik voor de verificatie van de authenticiteit van reisdocumenten en ondersteunende documentatie van natuurlijke personen en de opsporing van niet-authentieke documenten door de controle van hun beveiligingskenmerken;
  - Ondersteunen bij het onderzoek van asielaanvragen, aanvragen voor een visum en aanvragen voor een verblijfsvergunning, evenals gerelateerde klachten met betrekking tot de geschiktheid van de natuurlijke personen die een aanvraag voor een status indienen;
- **Rechtsbedeling en democratische processen:**
  - Ondersteuning van rechterlijke instanties bij het onderzoeken en uitleggen van feiten en de wet en bij de toepassing van het recht op een concrete reeks feiten.

Kenniscentrum
Data & Maatschappij

Vereisten & verplichtingen?

Vereisten en verplichtingen

Kenniscentrum Data & Maatschappij

Systeem voor risicobeheer

- Periodieke risicoanalyse
- Gehele levensduur AIS
- Maatregelen

Data en databeheer

- Passende databeheerspraktijken
- Gegevenskwaliteitsvereisten (*relevant, representatief, foutenvrij en volledig*)

Technische documentatie

- Doel: *controle faciliteren*
- Details in Annex IV (*relevantie parameters, testprocedures,…*)

Registratie

- Doel: *Traceerbaarheid werking AIS*
- Automatische registratie van events: "logging"
- Minimumvereisten voor biometrische identificatie

Vereisten en verplichtingen (2)

Kenniscentrum Data & Maatschappij

Transparantie en informatieverstrekking aan gebruikers

- Doel: *gebruikers in staat te stellen de output van het systeem te interpreteren en op passende wijze te gebruiken*
- Gebruiksaanwijzingen (met minimuminformatie)
  - *o.a. 'specificaties voor inputdata', mate van nauwkeurigheid,…*

Menselijk toezicht

- Doel: *doeltreffend menselijk toezicht toelaten*
- Ingebouwd ofwel door de gebruiker zelf
- Minimumvereisten:
  - *Werking monitoren*
  - *Automation bias*
  - *AI output negeren of ongedaan maken*
  - …

Nauwkeurigheid, robuustheid en cyberbeveiliging

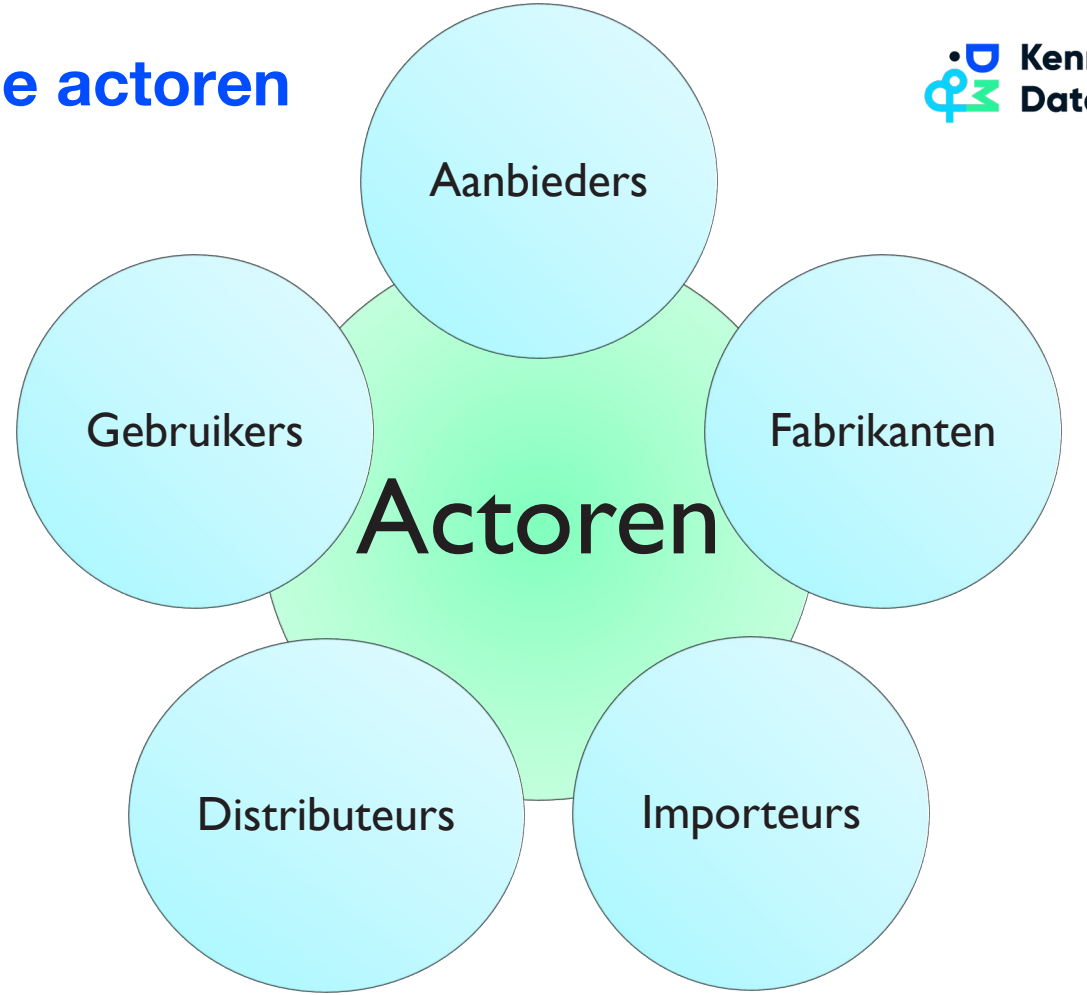
- Nauwkeurigheid: vermeld in GA
- Robuustheid: o.a. feedback loops beperken
- Cyberbeveiliging: o.a. tegen data poisoning en adversarial examples

Kenniscentrum
Data & Maatschappij

Rollen en verantwoordelijkheden?

Verschillende actoren

Kenniscentrum Data & Maatschappij



Aanbieders	Fabrikanten	Importeurs	Distributeurs	Gebruikers
Algemene nalevingsplicht	Algemene nalevingsplicht *(producten Annex II – Deel A + onder naam van F)*	Technische documentatie (+ gebruiksaanwijzingen)	Technische documentatie (+ gebruiksaanwijzingen)	Informatieplicht tav A & D autoriteiten *(bep. risico's + incidenten)*
Kwaliteitsbeheer *(prop. tav grootte organisatie)*		Conformiteitsbeoordeling	CE-markering	Retentie logs (mits controle)
Technische documentatie		CE-markering	Informatieplicht tav A,I & autoriteiten *(bep. risico's)*	
Retentie logs (mits controle)		Informatieplicht tav A & autoriteiten *(bep. risico's)*	Samenwerkingsplicht m. autoriteiten	
Conformiteitsbeoordeling		Samenwerkingsplicht m. autoriteiten	Corrigerende maatregelen *(of door A/I) (⇔ definitie?)*	
Registratie in HR DB		***Specifieke vpln***		
Corrigerende maatregelen *(info → D&I)*		Naam- en adresvermelding	Gepaste opslag- en vervoersomstandigheden	Gebruik & monitoring v werking // aanwijzingen
Informatieplicht tav autoriteiten *(bep. risico's)*		Gepaste opslag- en vervoersomstandigheden		Relevante inputdata (mits controle)
CE-markering				GEB (indien relevant)
Samenwerkingsplicht m. autoriteiten		*Weigering tot vermarkting bij niet-naleving vereisten*		

Art. 28 AIA: Wijziging van hoedanigheid

Kenniscentrum Data & Maatschappij

- Importeur/distributeur/gebruiker/derde ➔ aanbieder, indien:
  1. *Vermarkting AIS onder eigen (merk)naam*
  2. *Wijziging beoogde doel*
  3. *Ingrijpende wijziging AIS*

- **In geval 1**: Gezamenlijke verantwoordelijkheid?
- **In geval 2 en 3**: Initiële aanbieder ≠ aanbieder

Kenniscentrum
Data & Maatschappij

Next Steps?

Europees Parlement

Kenniscentrum Data & Maatschappij

- Situatie = vrij complex → verschillende comités hebben zeggenschap over delen van AIA
- **Belangrijkste actoren?**
  - Commissie Interne markt (Brando Benifei, S&D) &
  - Commissie Burgerlijke vrijheden, justitie en binnenlandse zaken (Dragos Tudorache, Renew Europe).
- **Anderen?**
  - Commissie Juridische zaken (Axel Voss, EVP) → transparantie & menselijk toezicht
  - Commissies Industrie, Cultuur en Onderwijs, Vervoer en Toerisme en Milieu
- **Resultaat?**
  - Vrij langzame vooruitgang *(+1000 wijzigingen*) - geconsolideerde positie tegen eind 2022?
- **Discussie?**
  - O.a. re definitie AI, verbod op biometrische identificatie in realtime,…

Raad van Ministers

Kenniscentrum Data & Maatschappij

- Gedeeltelijke compromissen onder SI en FR voorzitterschap
- O.a. re
  - Art. 1-7 & Annex I-III
  - Art. 8-15 & Annex IV
  - Art. 16-29
  - Art. 40-52
  - …

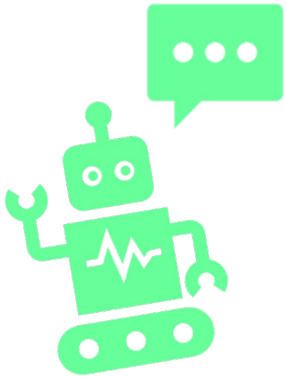
- Amendementen zijn vaak eerder klein maar talrijk

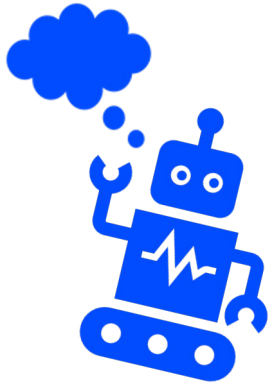
Wanneer?

Kenniscentrum
Data & Maatschappij

Kenniscentrum
Data & Maatschappij

Q&A





thomas.gils@kuleuven.be