Nederlands compliance Instituut

woensdag 11 december 2024

AI, Governance en Compliance

Start

Dit plaatje is gemaakt met ChatGPT op 27 november 2024

Nederlands compliance Instituut

woensdag 11 december 2024

AI, Governance en Compliance

Dit plaatje is gemaakt met ChatGPT op 27 november 2024

Even voorstellen: Eric Schuiling

- Eric Schuiling
- Kennis- en Programmamanager & Senior Compliance Adviseur

- Mobiel: 06 – 58 89 41 40
- Email: e.schuiling@compliance-instituut.nl

AI: (on)dankbaar onderwerp

“THE TOP-X AI TOOLS FOR PROFESSIONALS”

NotebookLM

?

Copilot

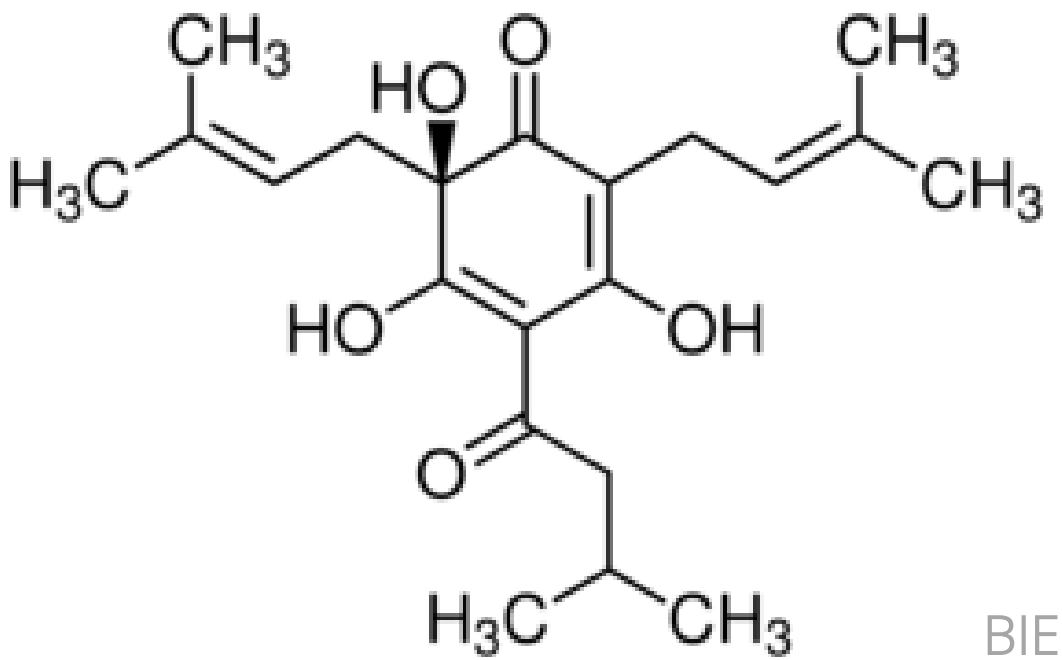
perplexity

YouLearn

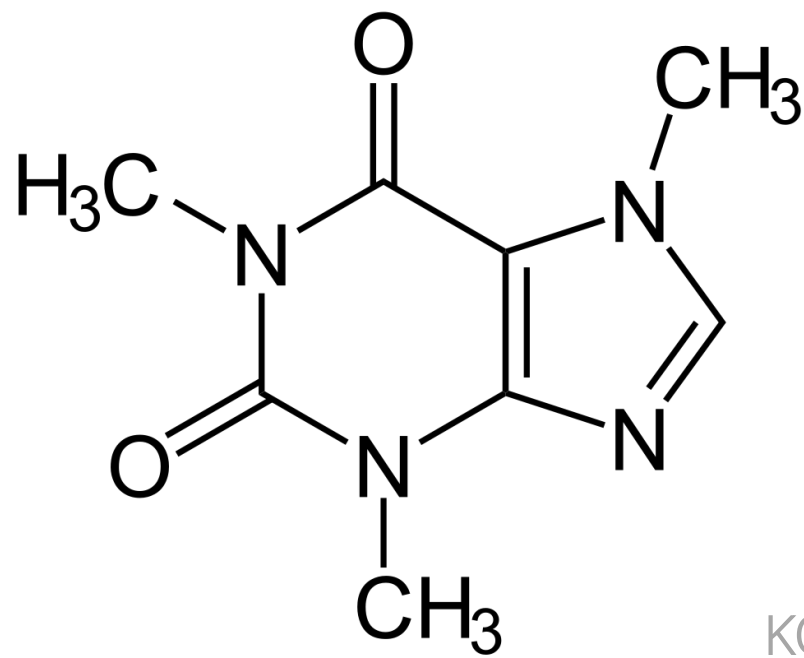
Hoe houden we het in de hand?

$$\left(\frac{\sin^2(x) + \cos^2(x)}{\int_0^\infty e^{-t}\,dt} \right)^{\sqrt{e^{2\pi i} \cdot \ln(e)}}$$

= 1



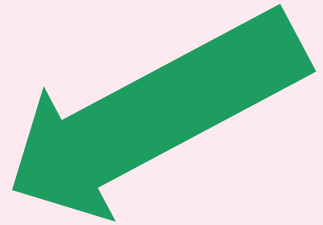
BIER



KOFFIE

ALIGNMENT-PROBLEEM

In welke mate zijn doelen en acties van AI-systemen in lijn met de waarden, doelstellingen en behoeften van de organisatie en haar stakeholders?



Rapportages vol geactualiseerde informatie en inzichten volgen elkaar in rap tempo op

AI: Goed of Niet goed?

Al die risico's...
Weet je wat? We verbieden het gewoon!

Bron: MT/Sprout

Shadow IT

De zon: Goed of Niet goed?

Inzicht

AI is er, net als de zon. Hoe je er mee omgaat bepaalt of het goed is of niet.

Wanneer stel je je vragen over de inzet van de AI-app?

A. Bij de PRESENTATIE van de AI-app door een collega
B. VOOR de bedrijfsbrede inzet van de AI-app
C. TIJDENS de bedrijfsbrede inzet van de AI-app
D. Tijdens de EVALUATIE van de inzet van de AI-app

van de inzet van de AI-app

Waar moeten je vragen over gaan?

AI: Kansen, maar zeker ook risico's genoeg

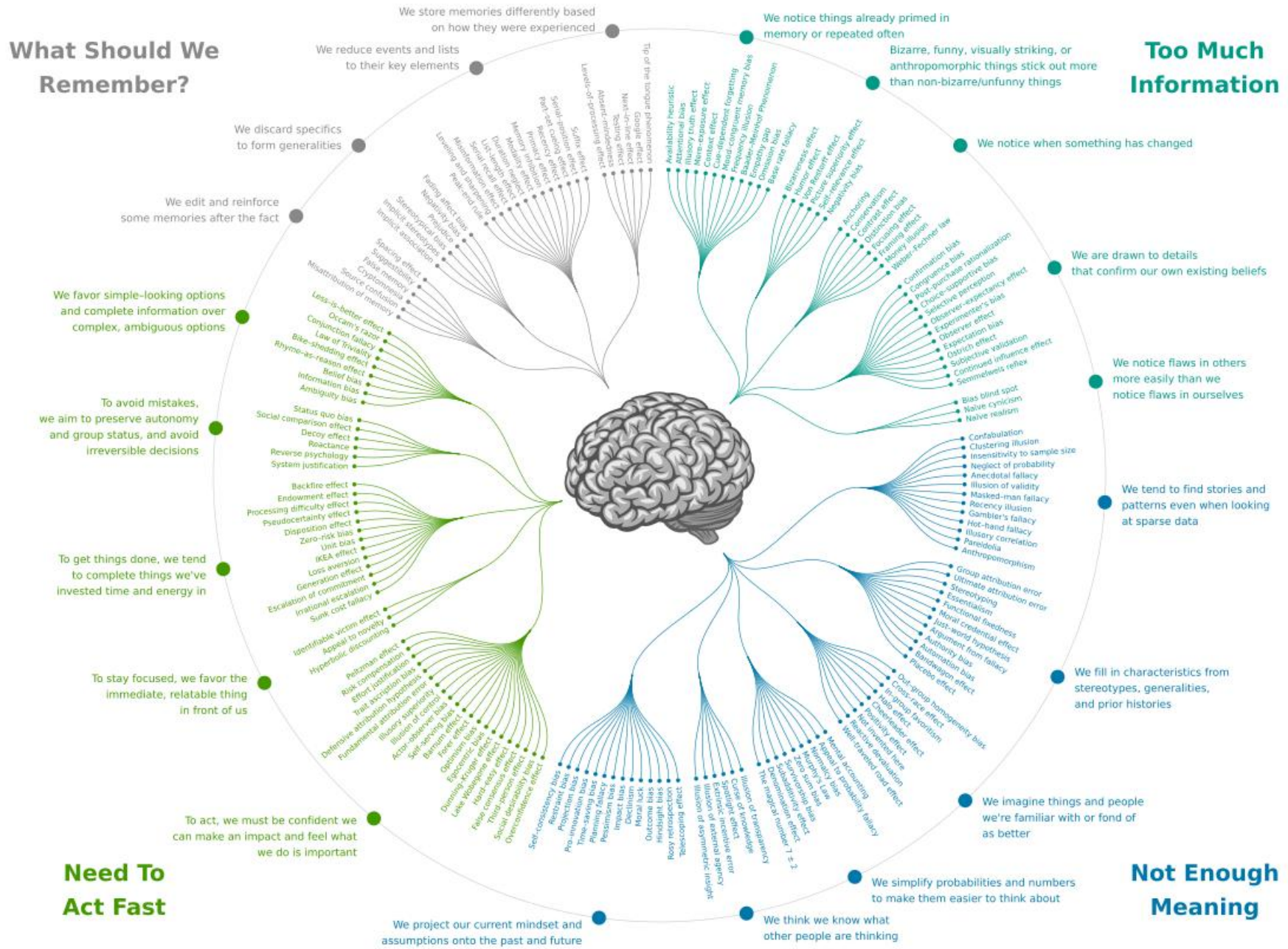
- Gegevensoverdrachtsrisico's
- (Gegevens-)veiligheidsrisico's
- Privacyrisico's
- Discriminatierisico's
- Fault Tolerance Risks
- Cognitieve gedragsmanipulatierisico's
- Risico van bevooroordeling (biases)
- Transparantierisico's
- Juridische risico's (bijv. auteursrechten)
- Financiële risico's
- Ethische risico's
- Verantwoordelijkheidsrisico's (autonomie)
- Solidariteitsrisico's (denk aan verzekeringen)
- Risico op verlies van menselijke controle
- Toezichtsrisico's
- Maatschappelijke risico's (bijv. werkloosheid)
- Duurzaamheidsrisico's
- ...

Bron: Microsoft Bing Copilot

Drie voorbeelden van Biases (Vooroordelen - Vooringenomenheden)

Data-bias	Algoritme-bias	Bevestigings-bias
Dit treedt op wanneer de **trainingsdata niet representatief is** voor de werkelijke populatie. Als bepaalde groepen ondervertegenwoordigd zijn in de dataset, kan het model vooringenomen zijn ten opzichte van die groepen. Bijvoorbeeld, als een gezichtsherkenningssysteem voornamelijk is getraind op afbeeldingen van lichte huidskleur, kan het moeite hebben met het nauwkeurig herkennen van donkere huidskleuren.	Dit ontstaat wanneer het model **zelf vooringenomenheden introduceert** tijdens het leren. Het kan gebeuren als het algoritme bepaalde kenmerken of patronen overmatig benadrukt, wat kan leiden tot ongewenste resultaten. Bijvoorbeeld, een AI-tool voor werving die onbedoeld mannen bevoordeelt bij het selecteren van kandidaten.	Dit type bias ontstaat door **onvolledige of subjectieve gegevens**. Het model kan bevooroordeelde conclusies trekken op basis van de beschikbare informatie. Bijvoorbeeld, als een aanbevelingssysteem alleen eerdere aankopen van een gebruiker in overweging neemt, kan het de gebruiker in een beperkt interessegebied houden en nieuwe mogelijkheden negeren.

THE COGNITIVE BIAS CODEX

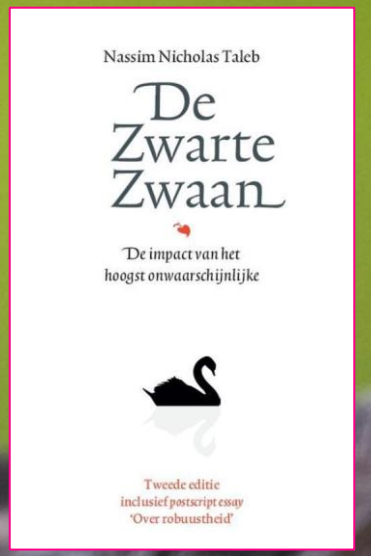


Bedenk goed:

=

OP DE COMPUTER
VAN IEMAND ANDERS

Risico van de Zwarte Zwaan

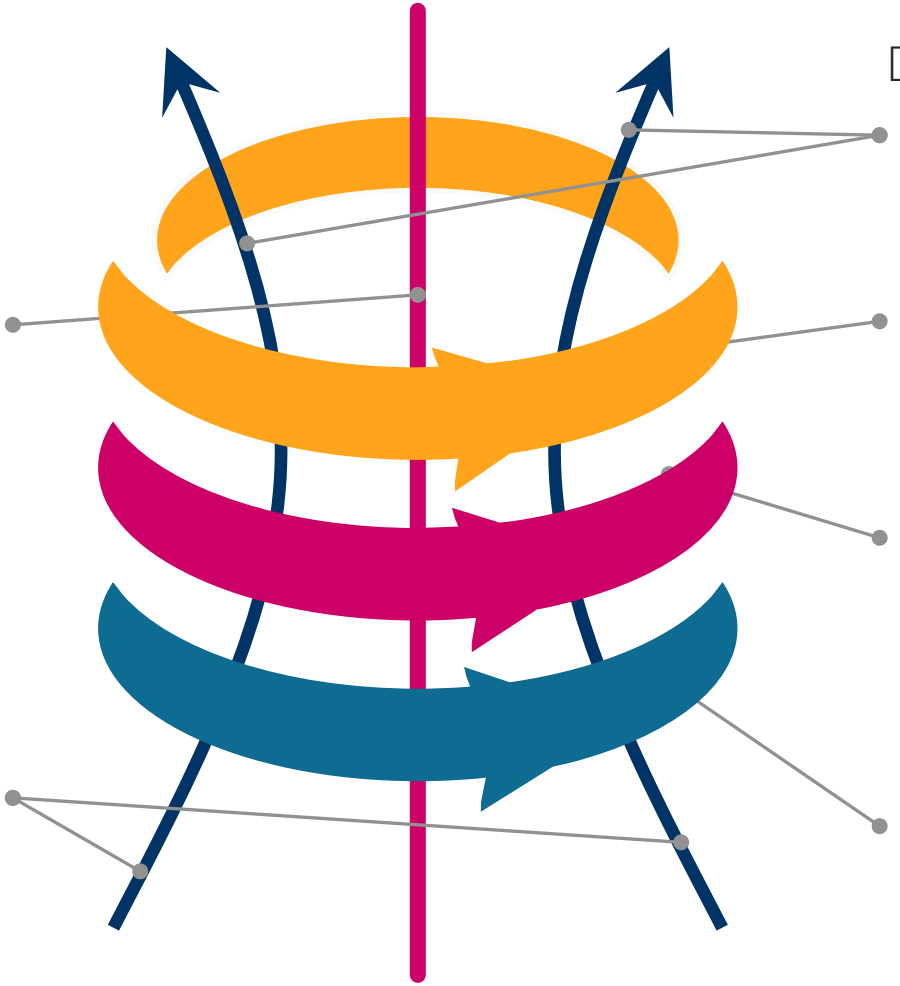


Bron: Nederlands Compliance Instituut

Het Twistermodel®: Groeien in compliance en integriteit

De rode lijn waar alles om draait: VERTROUWEN

De **kracht** van compliance, integriteit en ethiek met een motiverende werking op anderen



Door te **leren** en te **verbeteren** kunnen organisaties zich ontwikkelen

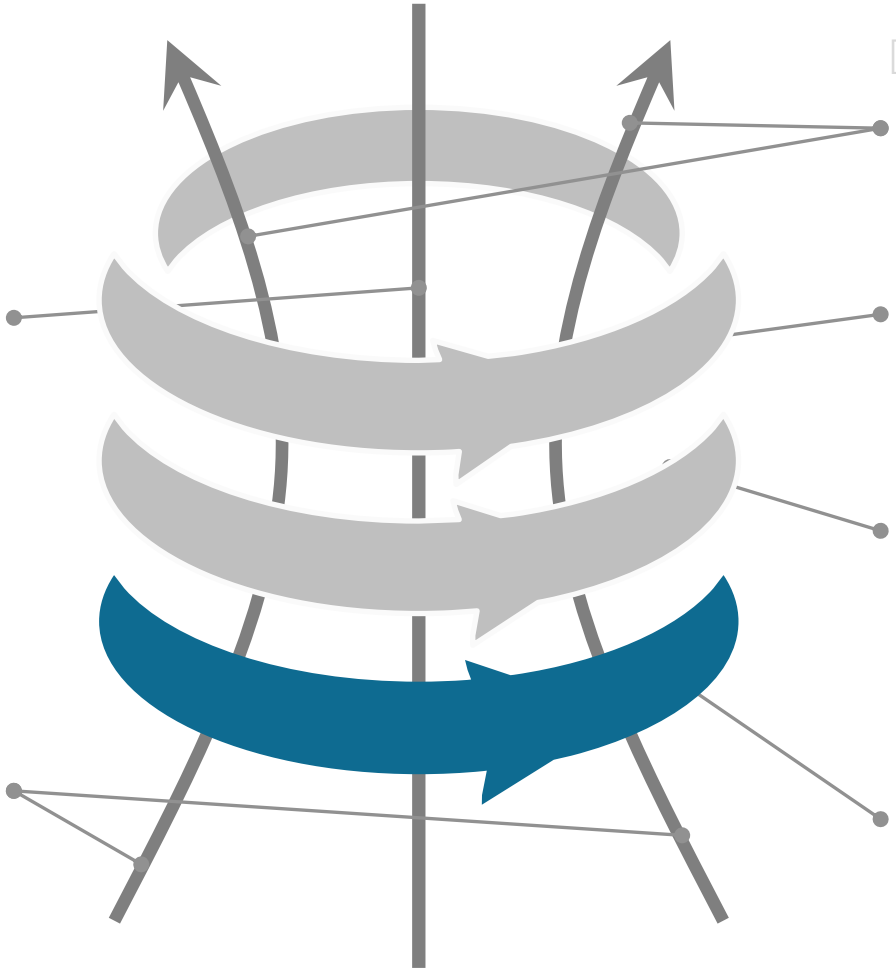
Het derde niveau: ETHIEK
Waardegedreven aanpak

Het tweede niveau: INTEGRITEIT
Beheersen van integriteitsrisico's

Het eerste niveau: COMPLIANCE
Voldoen aan wet- en regelgeving

Bron: Nederlands Compliance Instituut

Het Twistermodel®: Groeien in compliance en integriteit



De rode lijn waar alles om draait: VERTROUWEN

De **kracht** van compliance, integriteit en ethiek met een motiverende werking op anderen

Door te **leren** en te **verbeteren** kunnen organisaties zich ontwikkelen

Het derde niveau: ETHIEK
Waardegedreven aanpak

Het tweede niveau: INTEGRITEIT
Beheersen van integriteitsrisico's

Het eerste niveau: COMPLIANCE
Voldoen aan wet- en regelgeving

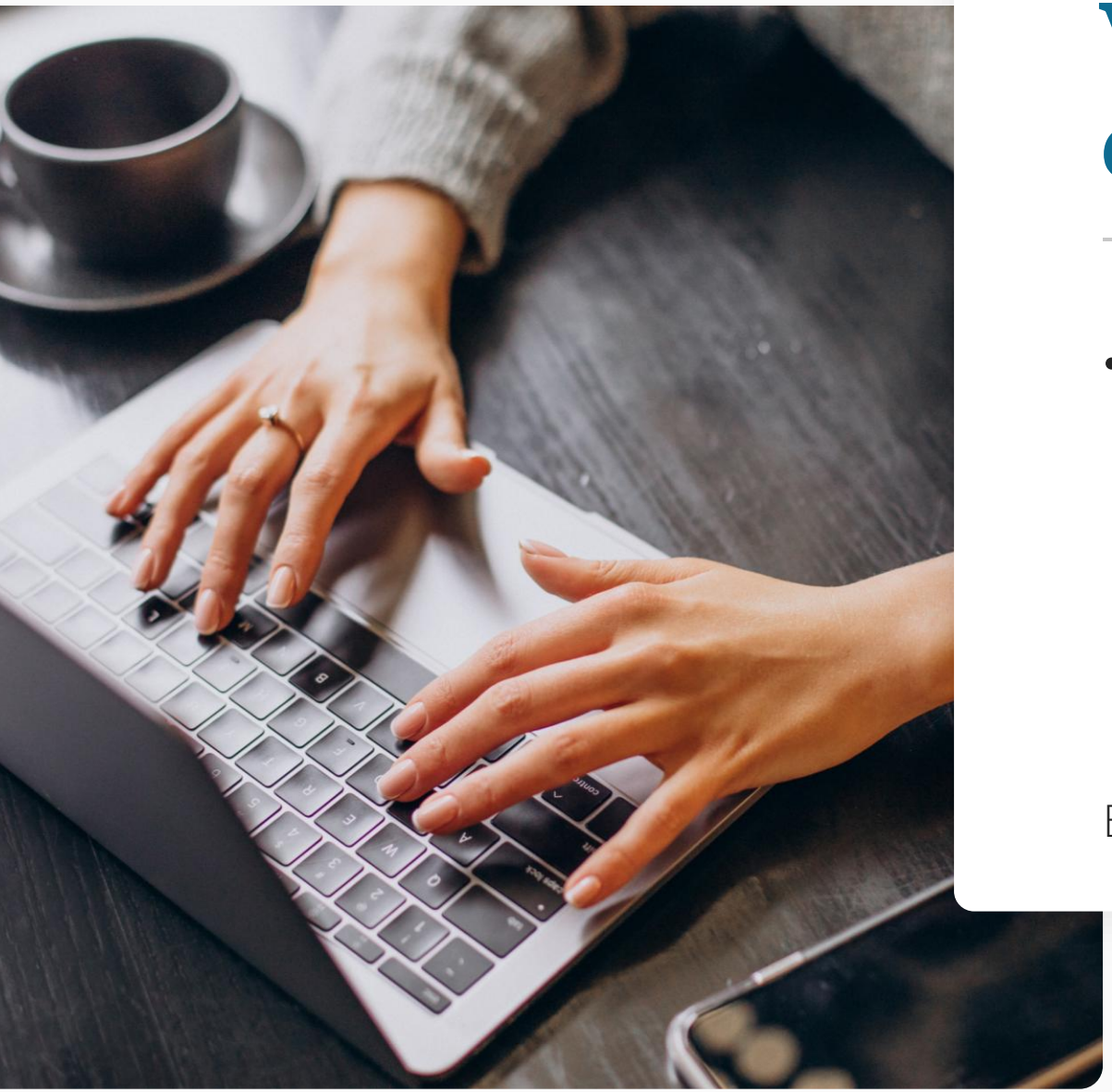
Wat zijn (o.a.) relevante wetten en regels mbt AI?

- Artikel 1 Grondwet
- EU Handvest van de grondrechten van de Europese Unie
- EU Verdrag voor de rechten van de mens (EVRM)
- EU Algemene Verordening Gegevensbescherming (AVG)
- EU AI Verordening
- EU Data Verordening
- EU Digital Operational Resilience Act (DORA)
- EU Network and Information Security Directive (NIS2)
- EU Critical Entities Resilience Directive (CER)

Artikel 1 Grondwet (gelijkheidsbeginsel) Waarom is dit nu juist zo van belang?

"Allen die zich in Nederland bevinden, worden in gelijke gevallen gelijk behandeld. Discriminatie wegens godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht of op welke grond dan ook, is niet toegestaan."


1 augustus 2024
De AI Verordening is van kracht.
Toepassing ervan
vindt in fasen plaats
gedurende
de komende jaren.
Bron: EU AI ImplementationTimeline

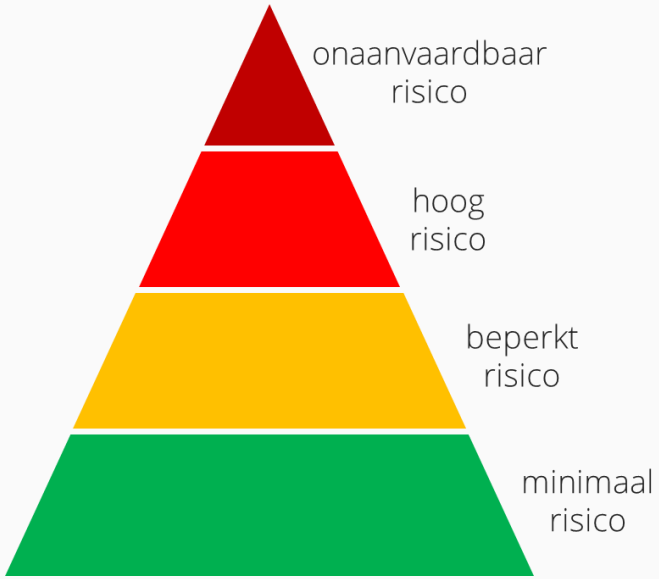


Voor wie geldt deze AI-verordening?

- De AI –verordening geldt voor:
  - **Ontwikkelaars, aanbieder en distributeurs** van AI-modellen en -systemen
  - Organisaties en bedrijven die de AI-systemen **gebruiken**.

Bron: Artikel 2 AIV en Digitale Overheid

EU AI Verordening



- Het is de **eerste** wet in haar soort en kan wereldwijd de norm voor AI-regelgeving worden.
- “Deze verordening is verbindend in al haar onderdelen en is **rechtstreeks toepasselijk** in elke lidstaat.
- De AI-verordening:
  - omvat **eisen en kaders voor de ontwikkeling en het gebruik** van AI-systemen door overheden en marktpartijen.
  - is bedoeld om innovatie en economische ontwikkelingen de ruimte te geven terwijl **publieke waarden worden beschermd**
- **Risicogebaseerde-aanpak**: hoe groter het risico is op schade voor de samenleving, des te strenger de regels zijn.

Bron: Uitleg EU AI Act en Nieuwsbericht Digitale Overheid

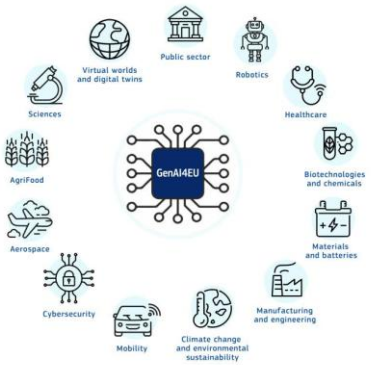
LET OP:
AI Geletterdheid (art 4)

- Aanbieders en gebruiksverantwoordelijken van AI-systemen nemen maatregelen om, zoveel als mogelijk, te zorgen voor een **toereikend niveau van AI-geletterdheid bij hun personeel** en andere personen die namens hen AI-systemen exploiteren en gebruiken, en **houden daarbij rekening met hun technische kennis, ervaring, onderwijs en opleiding en de context** waarin de AI-systemen zullen worden gebruikt, evenals met de personen of groepen personen ten aanzien van wie de AI-systemen zullen worden gebruikt.

Bron: Artikel 4 AIV

LET OP
Van toepassing vanaf 2 februari 2025

GenAI4EU



- EC lanceerde in januari 2024 het '[AI Innovation Package]':
  - GenAI4EU initiatief
  - EU AI Office
- Doel van het **GenAI4EU initiatief** is het ondersteunen van startups en het MKB bij het ontwikkelen van betrouwbare kunstmatige intelligentie conform de Europese waarden, wetten en regels
- 14 toepassingsgebieden, waaronder:
  - Robotica
  - Gezondheidszorg en Biotech
  - Overheid en Virtuele werelden

EU AI Bureau

- Opgericht 29 mei 2024
- Versterken van EU-leiderschap op het gebied van veilige en betrouwbare AI
- O.a. ook toezicht op AI Verordening in nauwe samenwerking met EU lidstaten (en hun toezichthouders)

Advies AI-toezichtsstructuur (AP/RDI – nov 2024)

- In de AI-verordening staan regels voor het verantwoord ontwikkelen en gebruiken van AI door bedrijven, overheden en andere organisaties. Goed georganiseerd toezicht hierop geeft vertrouwen aan consumenten en schept duidelijkheid voor organisaties en bedrijfsleven. Zij kunnen **blijven schakelen met de (sectorale) toezichthouders die zij al kennen**.
- Door de vele verschillende AI-toepassingen zijn er ook veel verschillende risico's. Een AI-systeem in speelgoed brengt andere risico's met zich mee dan bijvoorbeeld een AI-systeem voor werving en selectie. De RDI en de AP adviseren daarom om het toezicht op AI in de verschillende sectoren en domeinen **zoveel mogelijk te laten aansluiten bij het reguliere toezicht**.
- In het eindadvies over het toezicht op de AI-verordening worden **coördinerende rollen toebedeeld aan de RDI en de AP**. Vanuit een expertrol ondersteunen en adviseren zij andere toezichthouders en faciliteren ze de samenwerking.

Toezichthouders in Nederland (1/2)

AUTORITEIT
PERSOONSGEGEVENS

De AP is sinds 2023 de coördinerend toezichthouder voor algoritmes en AI. Ze spelen een centrale rol in het toezicht op AI-systemen en zijn de regisseur van de Algoritme en AI Kamer (AAK)

Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken en Klimaat

Samen met de AP benadrukt de RDI het belang van samenwerking tussen toezichthouders. Ze adviseren dat het toezicht op AI zoveel mogelijk moet aansluiten bij het reguliere toezicht dat al bestaat. De RDI is betrokken bij het reguleren van verantwoorde AI-toepassingen, zoals regulatory sandboxes

Nederlandse Voedsel- en Warenautoriteit
Ministerie van Landbouw, Natuur en Voedselkwaliteit

De NVWA blijft bijvoorbeeld speelgoed keuren, zelfs als er AI in verwerkt zit

Toezichthouders in Nederland (2/2)

Inspectie Gezondheidszorg en Jeugd
Ministerie van Volksgezondheid, Welzijn en Sport

De IGJ houdt toezicht op AI in medische apparaten

AFM

DeNederlandscheBank
EUROSYSTEEM

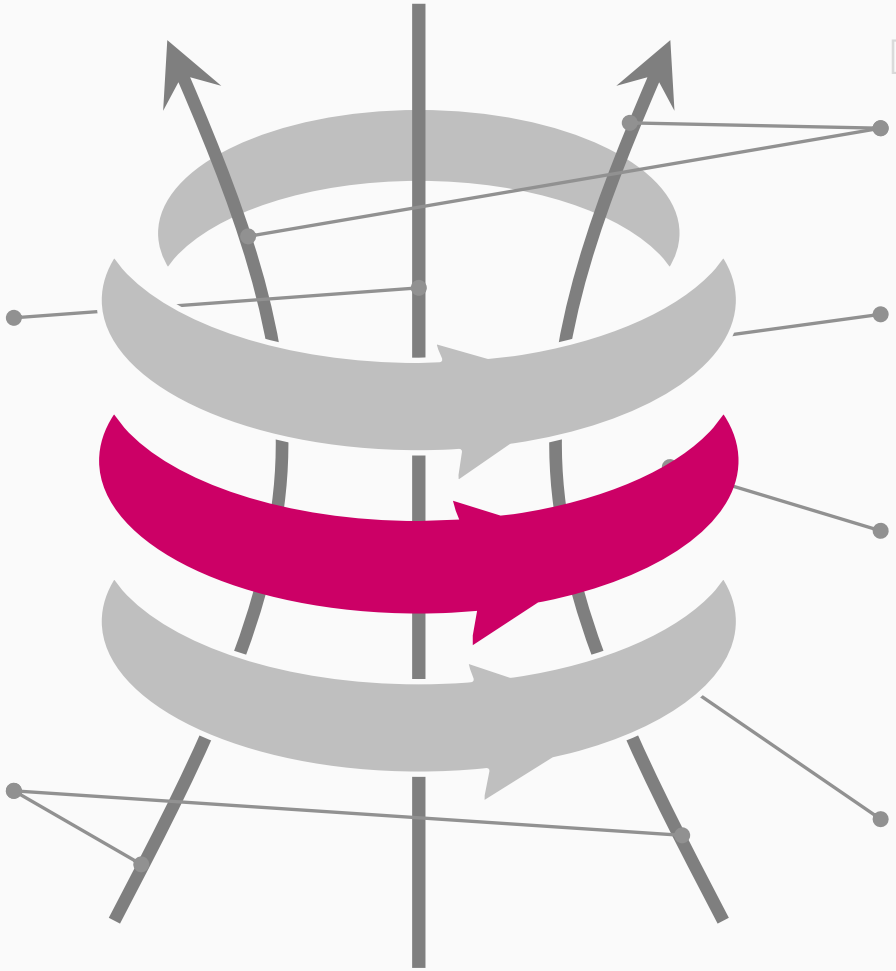
Deze toezichthouders zijn verantwoordelijk voor het markttoezicht op AI in de financiële sector

Inspectie Leefomgeving en Transport
Ministerie van Infrastructuur en Waterstaat

De ILT houdt toezicht op AI-systemen in de kritieke infrastructuur

Bron: Nederlands Compliance Instituut

Het Twistermodel®: Groeien in compliance en integriteit



De rode lijn waar alles om draait: VERTROUWEN

De **kracht** van compliance, integriteit en ethiek met een motiverende werking op anderen

Door te **leren** en te **verbeteren** kunnen organisaties zich ontwikkelen

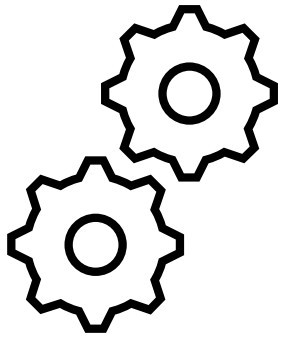
Het derde niveau: ETHIEK Waardegedreven aanpak

Het tweede niveau: INTEGRITEIT Beheersen van integriteitsrisico's

Het eerste niveau: COMPLIANCE Voldoen aan wet- en regelgeving

Is het zo simpel?

Vaak is het niet zo simpel: Drie modellen die je kunnen helpen

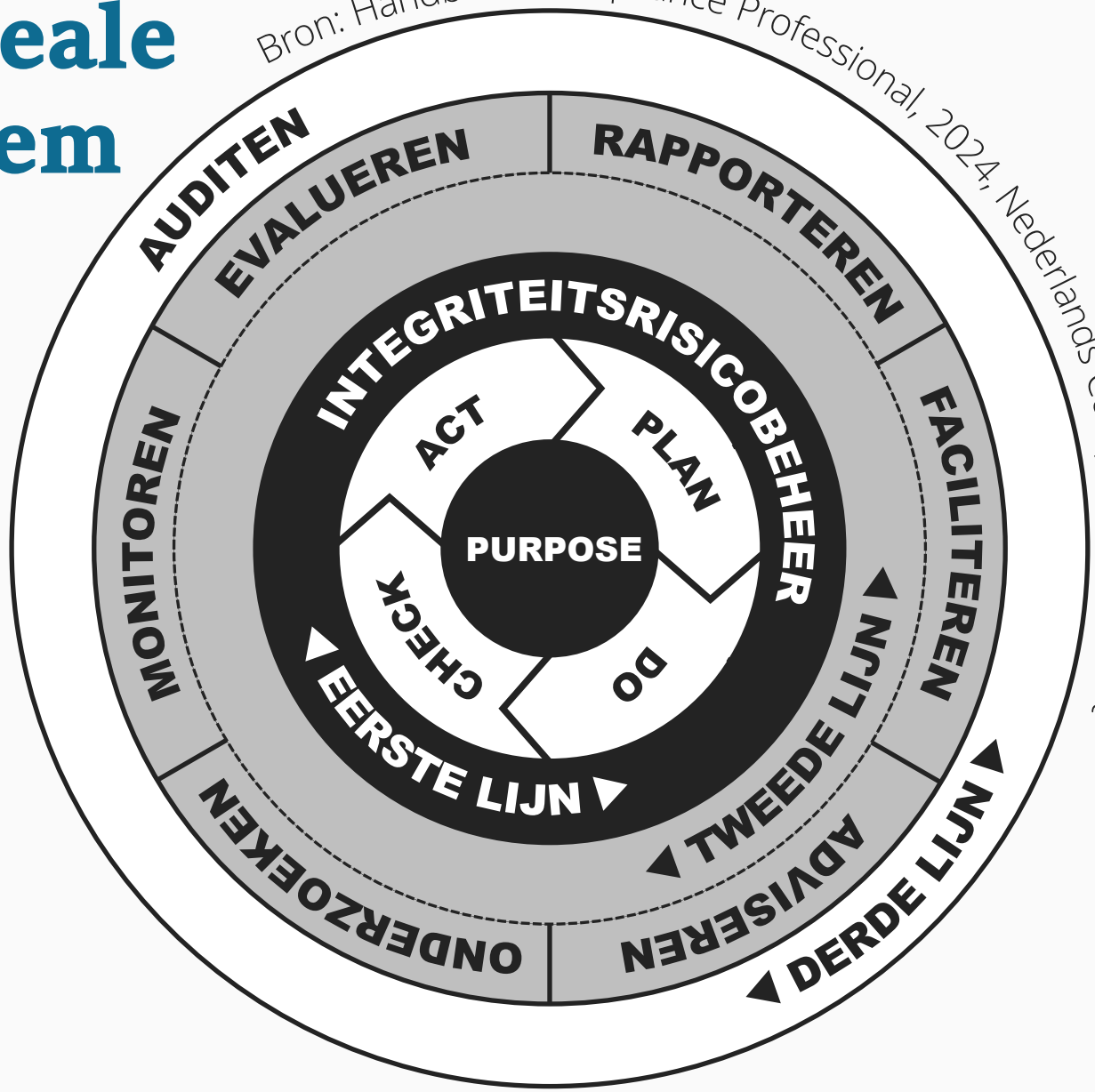


Het Compliance Systeem

Het SIRA-6-stappen-model

Het AP AI Governance model

Een blik op het ideale Compliance Systeem



Voorziet in:

- Eenvoud
- Purpose centraal
- Drie lijnen model IIA
- Verantwoordelijkheid
- Taken compliance officer(s)
- Tegenspraak (!)

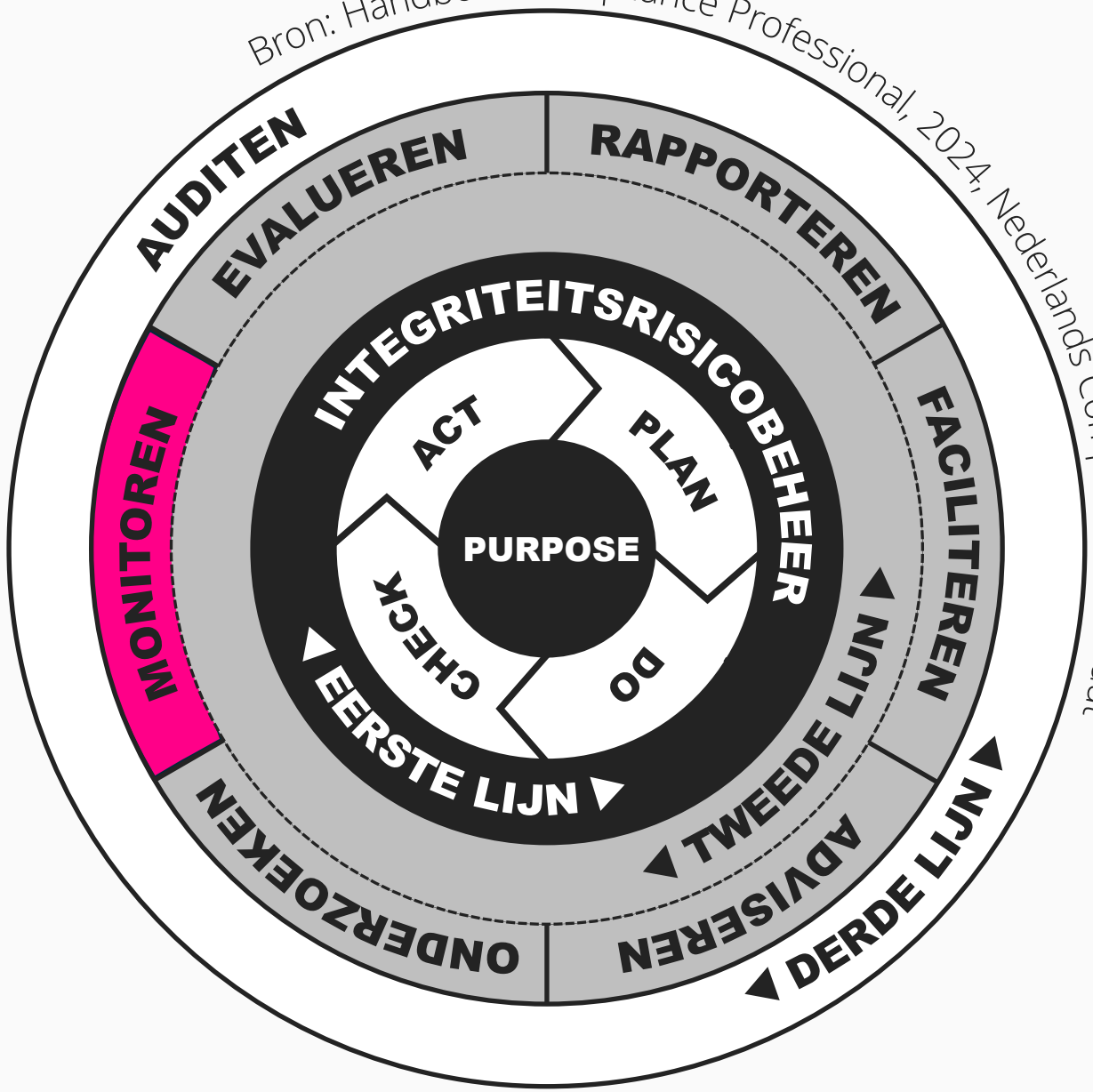
Leuk, maar HOE?

Algemene Rekenkamer

Toetsingskader algoritmes v2.0 (2023)

HLEG/AI

Assessment List for Trustworthy Artificial Intelligence (ALTAI) (2020)



Het SIRA-6-stappen-model
(Systematische Integriteitsrisicoanalyse)

- S ystematische
- I ntegriteits
- R isico-
- A nalyse

Hoe kunnen we INTEGRITEIT definiëren?

Mensen reflecteren steeds opnieuw kritisch en systematisch op hun kernverantwoordelijkheden en stellen zich voortdurend vragen stellen als:
Hoe doe ik mijn werk goed?
Doe ik recht aan de situatie?
Houd ik in voldoende mate rekening met de rechten, belangen en het welzijn van alle belanghebbenden?

ZORGVULDIG

Mensen kunnen aangeven hoe hun handelen past bij hun kernverantwoordelijkheden en kerntaken, bij de kernwaarden, regels, richtlijnen, wetten en andere bindende voorschriften van hun organisatie.

UITLEGBAAR

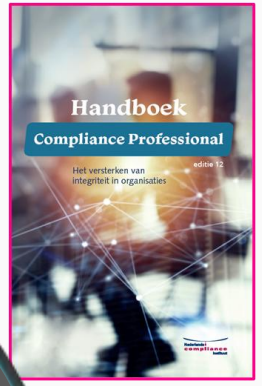
Mensen houden hun rug recht bij weerstanden en verleidingen;
Ze handelen niet onverantwoord omdat dit de weg van de minste weerstand is

STANDVASTIG

HANDELEN

Bron: Edgar Karssing

En wat kunnen we dan onder INEGRITEITSRISICO'S verst



De kans op het niet of onvoldoende naleven van wet- en regelgeving en het interne beleid

+

De kans op het niet of onvoldoende voldoen aan (gerechtvaardige) verwachtingen van stakeholders

Met ... ig...

Maatregelen v... autorit...

Aantasting van het vermogen

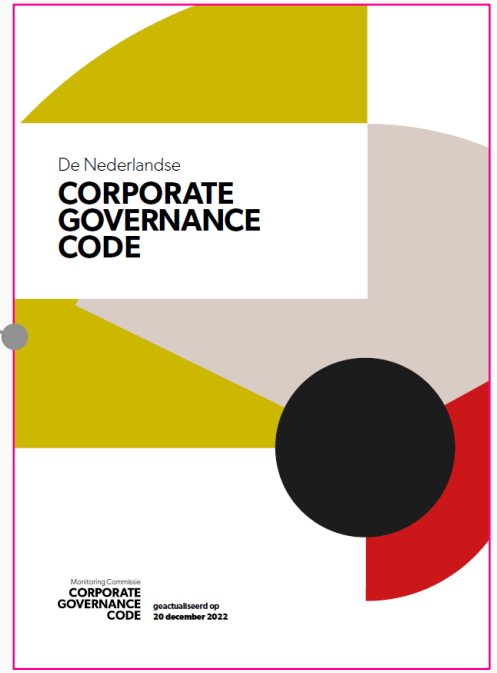
Aantasting van de reputatie van de organisatie

Wat zijn de gerechtvaardigde verwachtingen van u en uw organisatie?

HOOFDSTUK 1.
DUURZAME LANGE TERMIJN WAARDECREATIE

Principe 1.1 Duurzame lange termijn waardecreatie

Het bestuur is verantwoordelijk voor de continuïteit van de vennootschap en de met haar verbonden onderneming en de duurzame lange termijn waardecreatie van de vennootschap en de met haar verbonden onderneming. Het bestuur houdt rekening met de effecten van het handelen van de vennootschap en de met haar verbonden onderneming op mens en milieu en weegt daartoe de in aanmerking komende belangen van de *stakeholders*. De raad van commissarissen houdt toezicht op het bestuur ter zake.



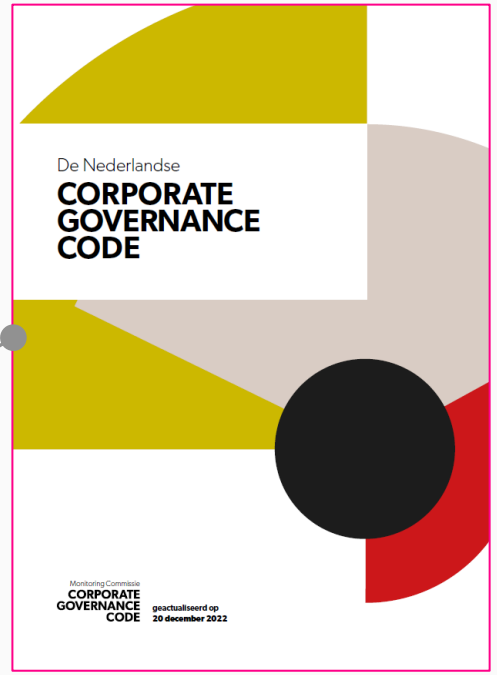
Wat zijn de gerechtvaardigde verwachtingen van u en uw organisatie?

1.1.1 Strategie voor duurzame lange termijn waardecreatie

Het bestuur ontwikkelt een visie op duurzame lange termijn waardecreatie van de vennootschap en de met haar verbonden onderneming en formuleert een daarbij passende strategie. Het bestuur formuleert concrete doelstellingen ter zake. Afhankelijk van de marktdynamiek kunnen korte termijn aanpassingen van de strategie nodig zijn.

Bij het vormgeven van de strategie wordt in ieder geval aandacht besteed aan:

i. de implementatie en haalbaarheid van de strategie;
ii. het door de vennootschap gevolgde bedrijfsmodel en de markt waarin de vennootschap en de met haar verbonden onderneming opereren;
iii. kansen en risico's voor de vennootschap;
iv. de operationele en financiële doelen van de vennootschap en de invloed ervan op de toekomstige positie in relevante markten;
v. de belangen van de *stakeholders*;
vi. de impact van de vennootschap en de met haar verbonden onderneming op het gebied van duurzaamheid, daaronder begrepen de effecten op mens en milieu;
vii. het leveren van een billijke bijdrage (*fair share*) aan de landen waarin de vennootschap opereert door betaling van belastingen; en
viii. de impact van nieuwe technologieën en veranderende businessmodellen.



Bron: DNB

Het SIRA-6-stappen-model (Systematische Integriteitsrisicoanalyse)

INHERENT RISK (BRUTO)

HOE IS HET BRUTO RISICO VAN DE ORGANISATIE TE OMSCHRIJVEN?

Let op: EEN BRUTO RISICO IS NIET HET FALEN VAN EEN CONTROL

1

RISK APPETITE

VALT HET BRUTO / INHERENT RISICO AL BINNEN DE RISK APPETITE VAN DE ORGANISATIE?

Zo ja: VERDERE MAATREGELEN ZIJN NIET NOODZAKELIJK

2

CONTROLS (BEHEERS-MAATREGELEN)

WELKE MAATREGELEN NEMEN WE OM HET BRUTO RISICO TE BEHEERSEN?

Let op: CONTROL is iets anders dan CONTROLE!

3

RESIDUAL RISK (NETTO)

OP BASIS VAN DE EFFECTIVITEIT VAN DE MAATREGELEN: WELK RISICO BLIJFT ER OVER?

4

RISK APPETITE

VALT HET NETTO / RESIDUAL RISICO BINNEN DE RISK APPETITE VAN DE ORGANISATIE?

Zo ja: VERDERE MAATREGELEN ZIJN NIET NOODZAKELIJK

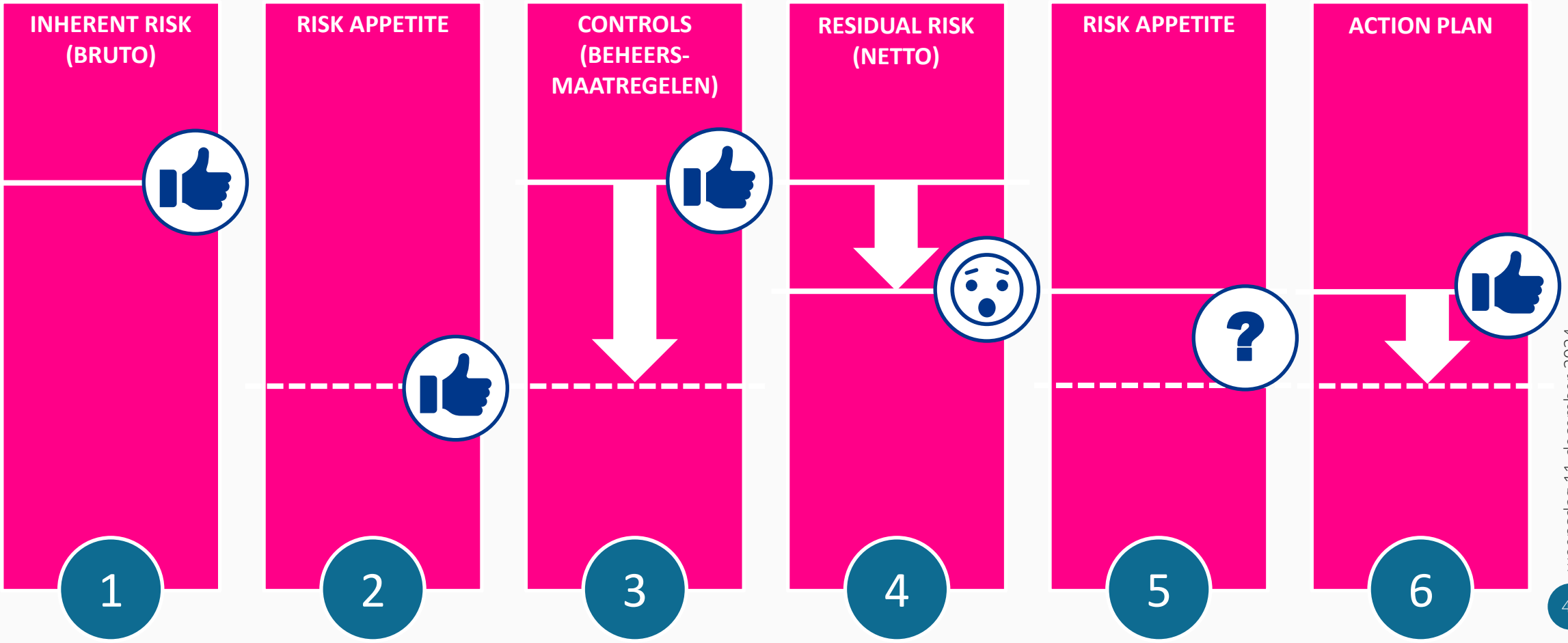
5

ACTION PLAN

WELKE RESPONS GEVEN WE OP BASIS VAN DE VERGELIJKING TUSSEN HET NETTO RISICO EN DE RISK APPETITE?

6

Het SIRA-6-stappen-model
(Systematische Integriteitsrisicoanalyse)



Het SIRA-6-stappen-model
(Systematische Integriteitsrisicoanalyse)

2015: "Meer waar dat moet, minder waar dat kan"

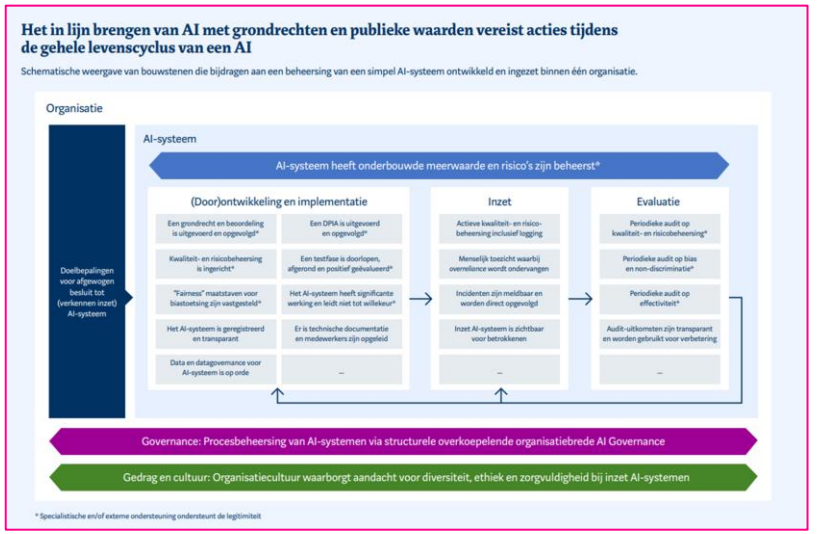
ZELF NADENKEN

2024: "Meer eigen reflectie, minder rigide"

Wat maakt het beheersen van AI-risico's zo complex?

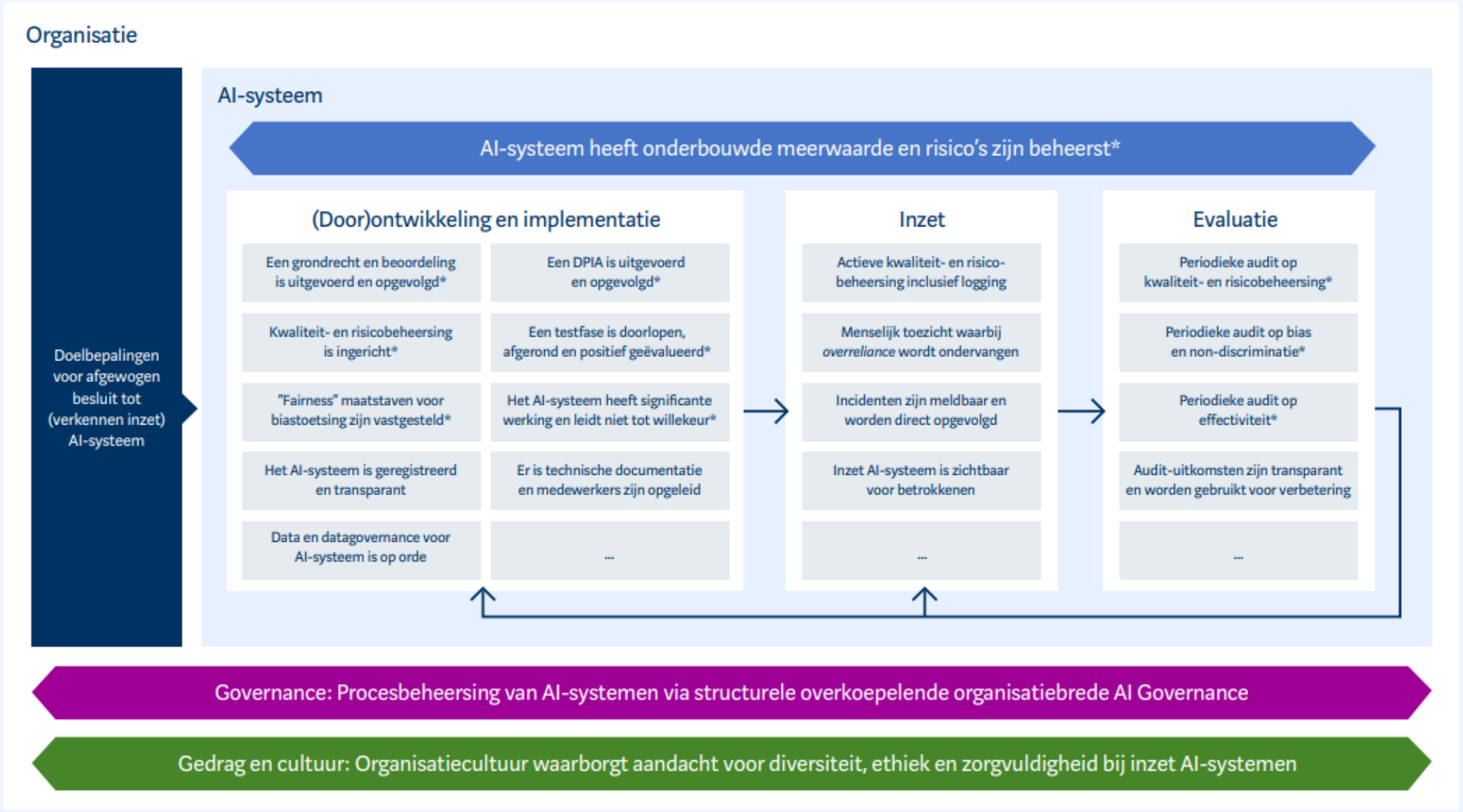
Let op HR-AI: art 9 AIV

- Er is **geen silver bullet** voor het beheersen van de risico's van AI-systemen.
- De schets in deze bijlage geeft op hoofdlijnen een overzicht van de samenhang in de risicobeheersing van een simpel AI-systeem



Het in lijn brengen van AI met grondrechten en publieke waarden vereist acties tijdens de gehele levenscyclus van een AI

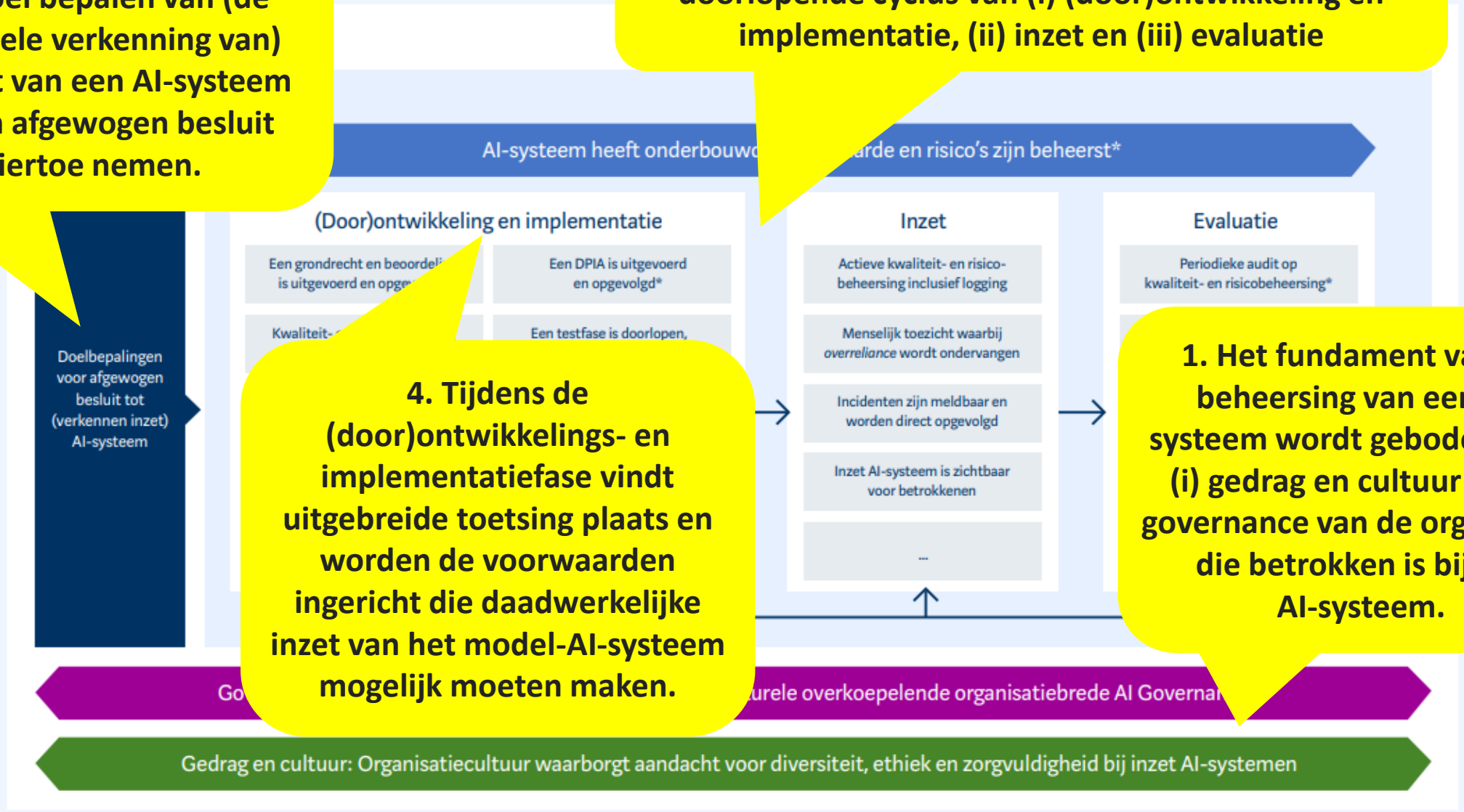
Schematische weergave van bouwstenen die bijdragen aan een beheersing van een simpel AI-systeem ontwikkeld en ingezet binnen één organisatie.



* Specialistische en/of externe ondersteuning ondersteunt de legitimiteit

2. Op organisatieniveau geldt vervolgens als eerste stap het doel bepalen van (de eventuele verkenning van) de inzet van een AI-systeem en een afgewogen besluit hiertoe nemen.

3. De continue risicobeheersing bij een AI-systeem binnen een organisatie vraagt vervolgens om een doorlopende cyclus van (i) (door)ontwikkeling en implementatie, (ii) inzet en (iii) evaluatie



4. Tijdens de (door)ontwikkelings- en implementatiefase vindt uitgebreide toetsing plaats en worden de voorwaarden ingericht die daadwerkelijke inzet van het model-AI-systeem mogelijk moeten maken.

1. Het fundament van de beheersing van een AI-systeem wordt geboden door (i) gedrag en cultuur en (ii) governance van de organisatie die betrokken is bij het AI-systeem.

* Specialistische en/of externe ondersteuning ondersteunt de legitimiteit

Het in lijn brengen van AI met grondrechten en publieke waarden vereist acties tijdens de gehele levenscyclus van een AI

Schematische weergave van bouwstenen die bijdragen aan een beheersing van

7. De periodieke evaluatie van het AI-systeem waarborgt dat het systeem doelmatig en kwalitatief in orde is en dat grondrechtenrisico's worden verkleind

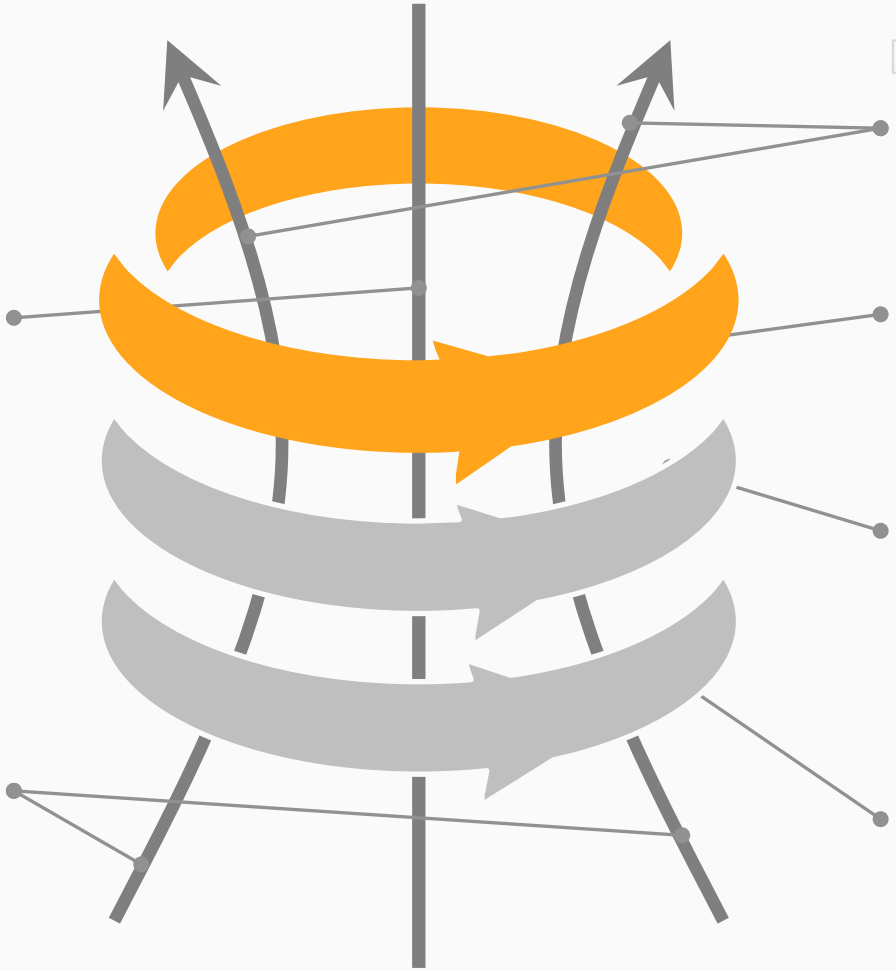
5. Ook moet er aandacht zijn voor testen, documentatie en registratie bij de ontwikkeling van AI.

6. Tijdens de inzetfase moeten gebruikers en derden in staat zijn om gevaren van en incidenten met een AI-systeem snel te identificeren en te corrigeren.

* Specialistische en/of externe ondersteuning ondersteunt de legitimiteit

Bron: Nederlands Compliance Instituut

Het Twistermodel®: Groeien in compliance en integriteit



De rode lijn waar alles om draait: VERTROUWEN

De **kracht** van compliance, integriteit en ethiek met een motiverende werking op anderen

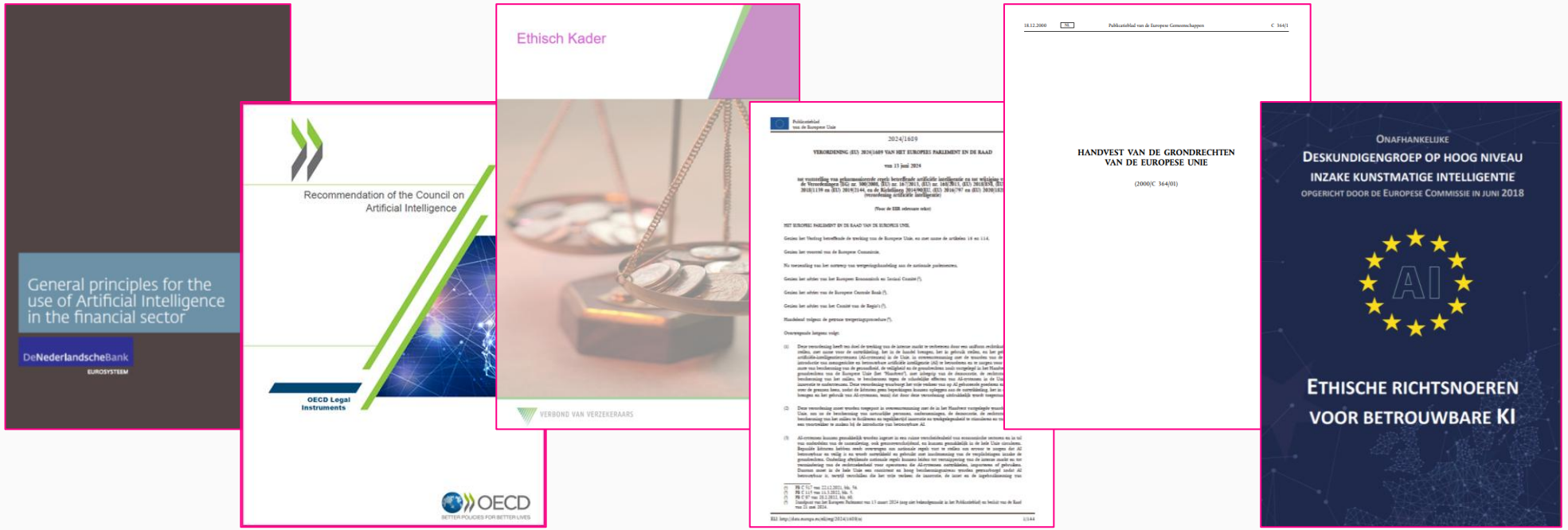
Door te **leren** en te **verbeteren** kunnen organisaties zich ontwikkelen

Het derde niveau:
ETHIEK
Waardegedreven aanpak

Het tweede niveau:
INTEGRITEIT
Beheersen van integriteitsrisico's

Het eerste niveau:
COMPLIANCE
Voldoen aan wet- en regelgeving

Een selectie van verschillende (ethische) kaders

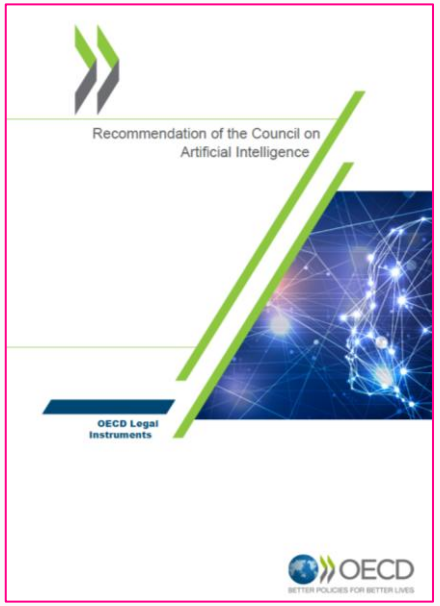


Verschillende kaders, waarden en principes

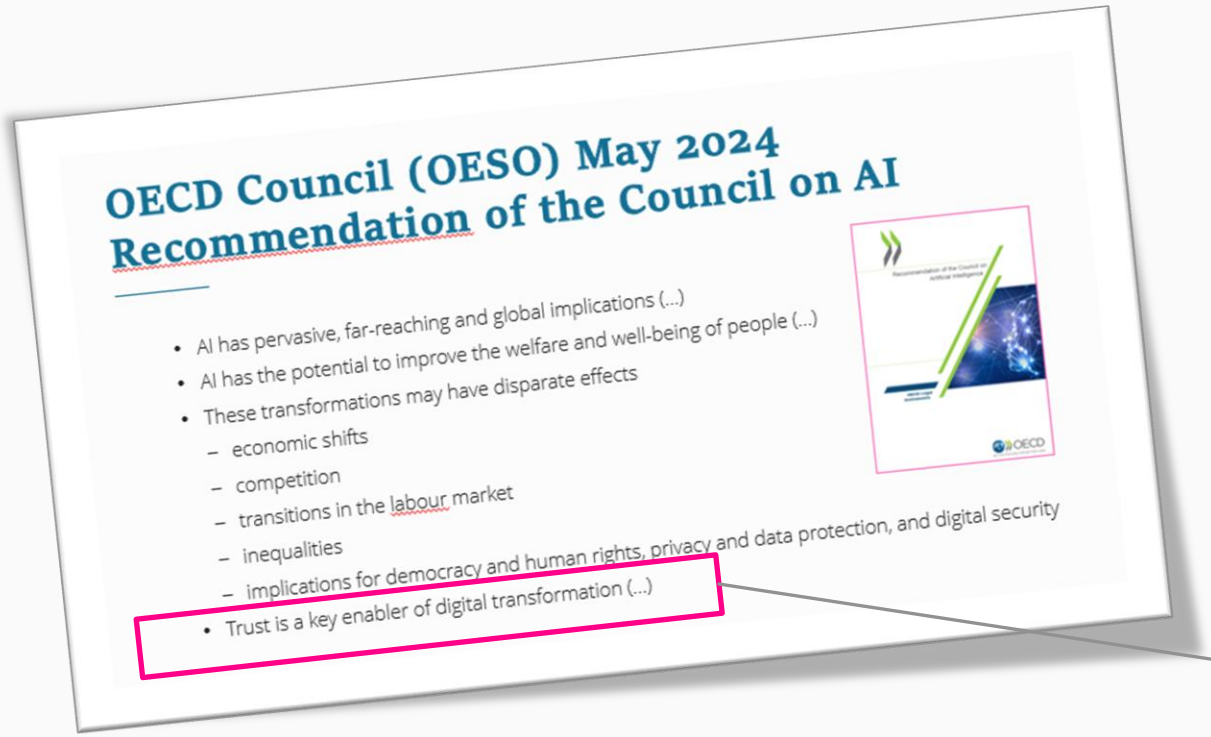
DNB	OESO	Verbond van Verzekeraars	AI Verordening *	EU Handvest Grondrechten	EU HLEG/AI
				Burgerschap	
Rechtvaardigheid			Bescherming van grondrechten	Rechtspleging	Eerlijkheid
Ethiek					
Vaardigheden					
		Autonomie / controle		Vrijheden	Menselijk handelen en toezicht
				Solidariteit	
	Inclusiviteit	Diversiteit / non-discriminatie		Gelijkheid	Diversiteit / non-discriminatie
		Privacy / data governance			Privacy/databeheer
	Respect		Menselijke waardigheid	Waardigheid	
			Gezondheid		
		Maatschappelijk welzijn			Milieu- en maatschappelijk welzijn
			Innovatie		
			Duurzaamheid		
Transparantie	Transparantie	Transparantie	Transparantie		Transparantie
	Uitlegbaarheid				
Degelijkheid	Robuustheid	Robuustheid			Robuustheid
	Veiligheid / beveiliging	Veiligheid	Veiligheid		Veiligheid
Verantwoordelijkheid	Verantwoordelijkheid	Verantwoording	Verantwoording		Verantwoordelijkheid

OECD Council (OESO) May 2024 Recommendation of the Council on AI

- AI has **pervasive, far-reaching and global implications** (…)
- AI has the **potential** to improve the welfare and well-being of people (…)
- These transformations may have **disparate effects**
  - economic shifts
  - competition
  - transitions in the labour market
  - inequalities
  - implications for democracy and human rights, privacy and data protection, and digital security
- **Trust** is a key enabler of digital transformation (…)



Moeten, willen en ‘moeten willen’



VERTROUWEN is het sleutelwoord als het gaat om een succesvolle digitale transformatie

Trust is a key enabler of digital transformation (...)

VRAAG

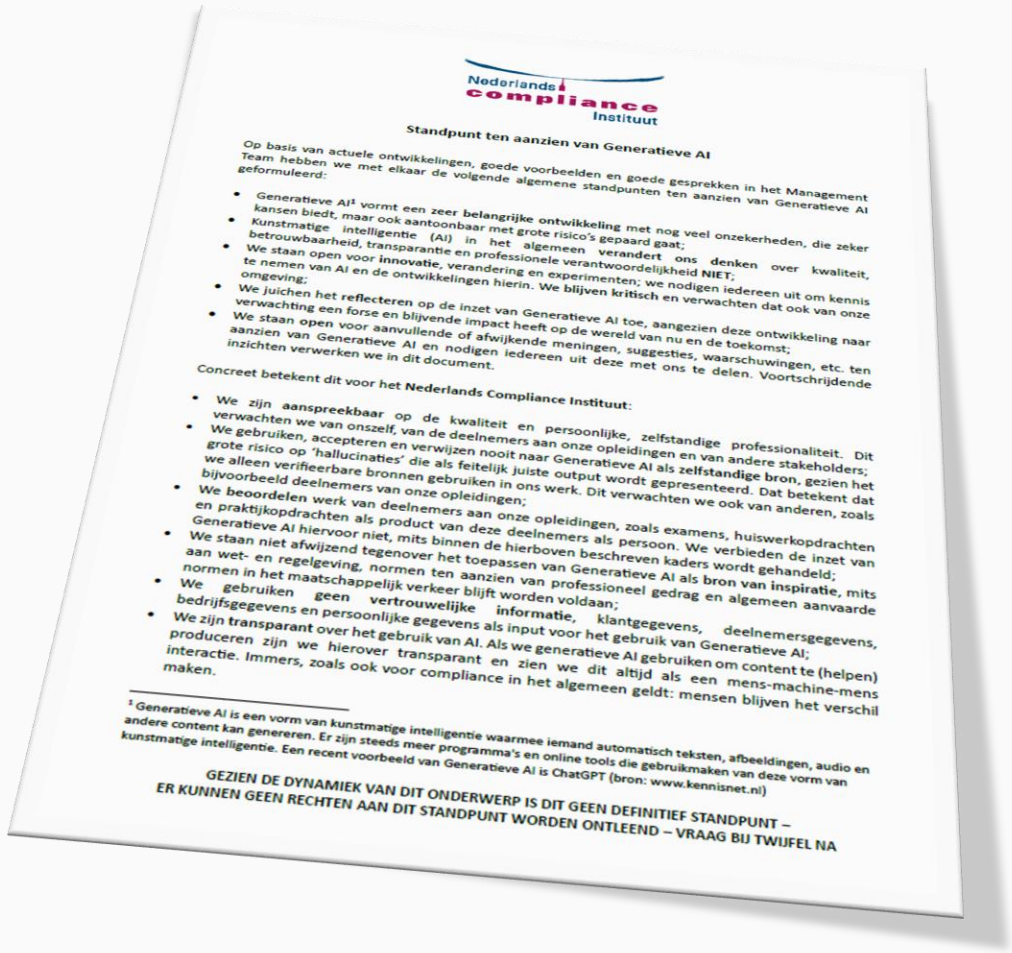
Hoe belangrijk is vertrouwen voor jouw organisatie?

Alles draait om …

Vijf praktische tips voor morgen

Stimuleer het opstellen van een AI zienswijze of AI beleid voor je organisatie

- Hoe kijkt onze organisatie tegen (generatieve) AI aan?
- Publiceer dit document op de interne en/of externe website en ga er met (interne en externe) stakeholders over in gesprek
- Inspiratie:
  - Statement NCI
  - Zienswijze ANP



Nederlands Compliance Instituut

Standpunt ten aanzien van Generatieve AI

Op basis van actuele ontwikkelingen, goede voorbeelden en goede gesprekken in het Management Team hebben we met elkaar de volgende algemene standpunten ten aanzien van Generatieve AI geformuleerd:

- Generatieve AI[1] vormt een zeer belangrijke ontwikkeling met nog veel onzekerheden, die zeker kansen biedt, maar ook aantoonbaar met grote risico's gepaard gaat;
- Kunstmatige intelligentie (AI) in het algemeen verandert ons denken over kwaliteit, betrouwbaarheid, transparantie en professionele verantwoordelijkheid NIET;
- We staan open voor innovatie, verandering en experimenten; we nodigen iedereen uit om kennis te nemen van AI en de ontwikkelingen hierin. We blijven kritisch en verwachten dat ook van onze omgeving;
- We juichen het reflecteren op de inzet van Generatieve AI toe, aangezien deze ontwikkeling naar verwachting een forse en blijvende impact heeft op de wereld van nu en de toekomst;
- We staan open voor aanvullende of afwijkende meningen, suggesties, waarschuwingen, etc. ten aanzien van Generatieve AI en nodigen iedereen uit deze met ons te delen. Voortschrijdende inzichten verwerken we in dit document.

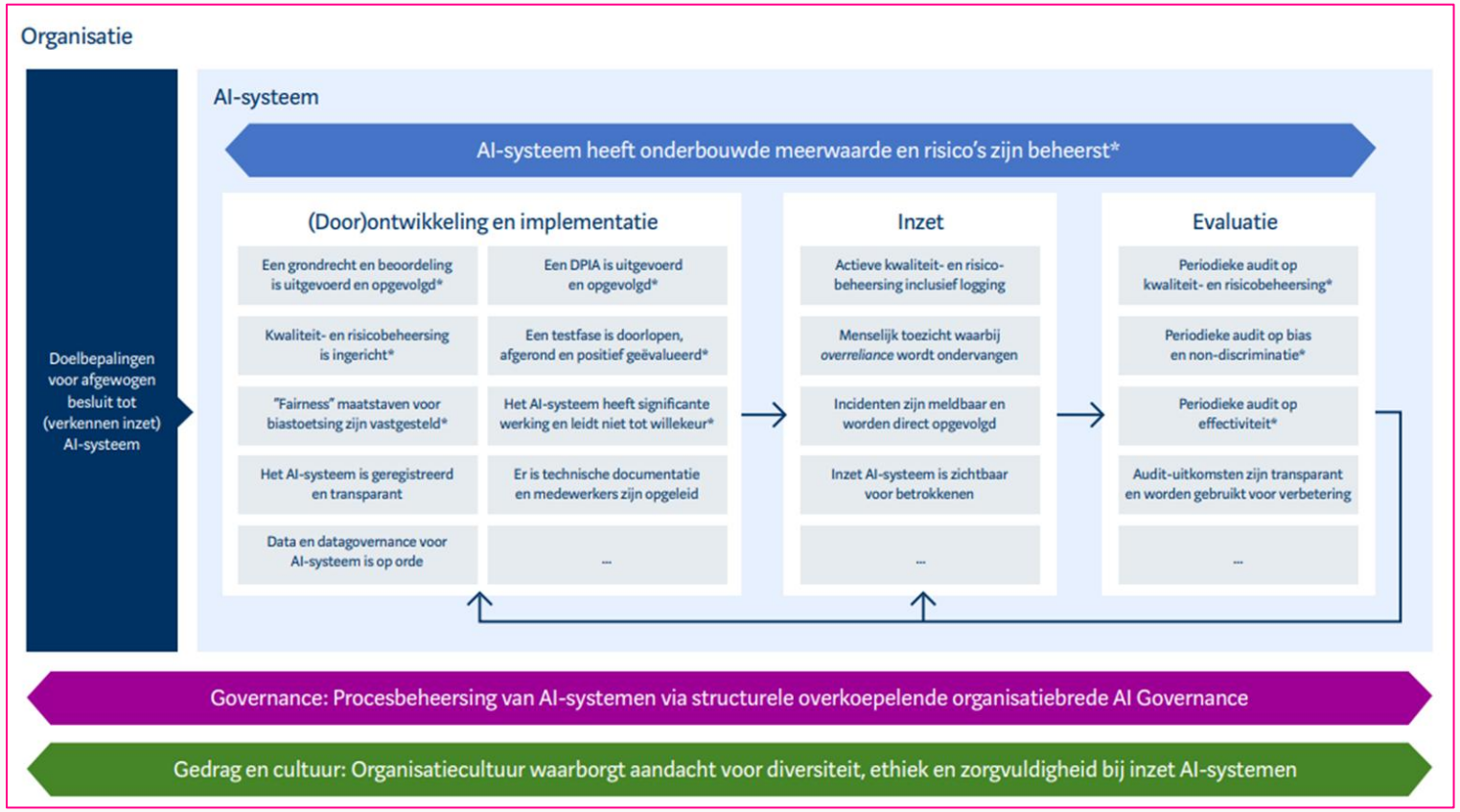
Concreet betekent dit voor het Nederlands Compliance Instituut:

- We zijn aanspreekbaar op de kwaliteit en persoonlijke, zelfstandige professionaliteit. Dit verwachten we van onszelf, van de deelnemers aan onze opleidingen en van andere stakeholders;
- We gebruiken, accepteren en verwijzen nooit naar Generatieve AI als zelfstandige bron, gezien het grote risico op 'hallucinaties' die als feitelijk juiste output wordt gepresenteerd. Dat betekent dat we alleen verifieerbare bronnen gebruiken in ons werk. Dit verwachten we ook van anderen, zoals bijvoorbeeld deelnemers van onze opleidingen;
- We beoordelen werk van deelnemers aan onze opleidingen, zoals examens, huiswerkopdrachten en praktijkopdrachten als product van deze deelnemers als persoon. We verbieden de inzet van Generatieve AI hiervoor niet, mits binnen de hierboven beschreven kaders wordt gehandeld;
- We staan niet afwijzend tegenover het toepassen van Generatieve AI als bron van inspiratie, mits aan wet- en regelgeving, normen ten aanzien van professioneel gedrag en algemeen aanvaarde normen in het maatschappelijk verkeer blijft worden voldaan;
- We gebruiken geen vertrouwelijke informatie, klantgegevens, deelnemersgegevens, bedrijfsgegevens en persoonlijke gegevens als input voor het gebruik van Generatieve AI;
- We zijn transparant over het gebruik van AI. Als we generatieve AI gebruiken om content te (helpen) produceren zijn we hierover transparant en zien we dit altijd als een mens-machine-mens interactie. Immers, zoals ook voor compliance in het algemeen geldt: mensen blijven het verschil maken.

[1] Generatieve AI is een vorm van kunstmatige intelligentie waarmee iemand automatisch teksten, afbeeldingen, audio en andere content kan genereren. Er zijn steeds meer programma's en online tools die gebruikmaken van deze vorm van kunstmatige intelligentie. Een recent voorbeeld van Generatieve AI is ChatGPT (bron: www.kennisnet.nl)

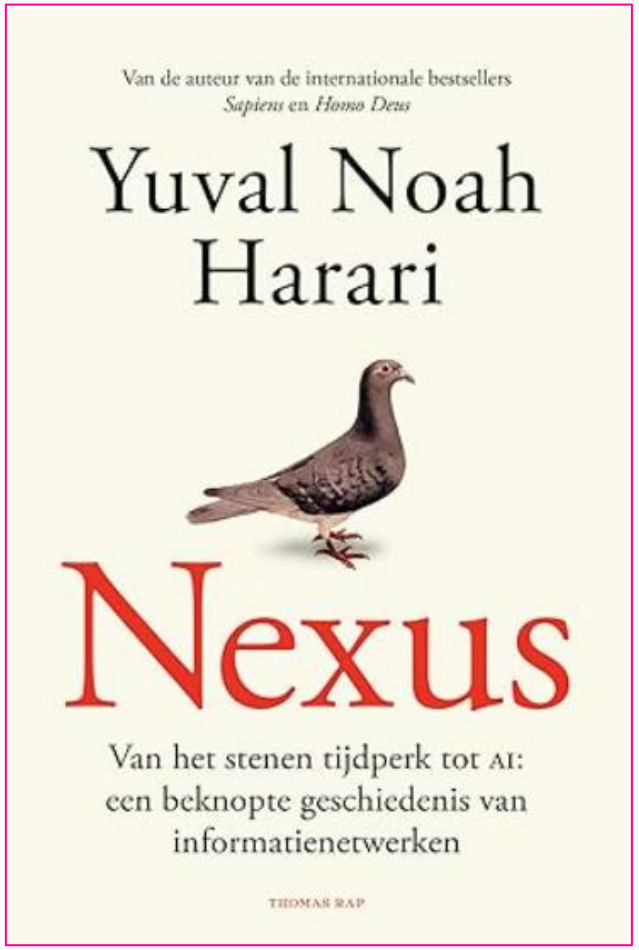
GEZIEN DE DYNAMIEK VAN DIT ONDERWERP IS DIT GEEN DEFINITIEF STANDPUNT – ER KUNNEN GEEN RECHTEN AAN DIT STANDPUNT WORDEN ONTLEEND – VRAAG BIJ TWIJFEL NA

Help de contouren van AI Governance vaststellen voor jouw organisatie



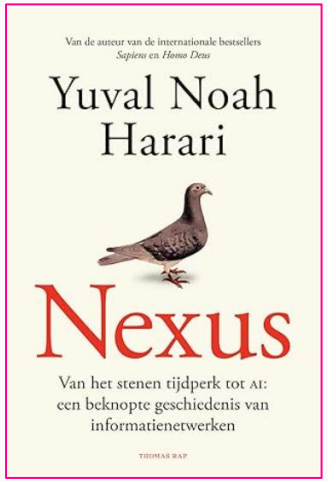
Zie verder: AI Governance (AP)

Lees het boek NEXUS van Yuval Noah Harari



"Van het stenen tijdperk tot AI: een beknopte geschiedenis van informatienetwerken"

Lees het boek NEXUS van Yuval Noah Harari



Uitdagingen:

- Manipulatie en desinformatie
- Verlies van privacy
- Machtsverschuiving naar AI
- Verstoring van de arbeidsmarkt

Kansen:

- Verbetering van de gezondheidszorg
- Voorkomen van ongelukken
- Gepersonaliseerd onderwijs
- Beter begrip van onszelf

Conclusie: AI heeft de potentie om de wereld op zowel positieve als negatieve manieren te veranderen. Het is essentieel dat we de ontwikkeling van AI zorgvuldig sturen en reguleren om de risico's te minimaliseren en de voordelen te maximaliseren. Harari benadrukt dat we ons moeten richten op het ontwikkelen van "levende instituten" die de uitdagingen van AI kunnen aanpakken en de technologie in de juiste richting kunnen sturen.

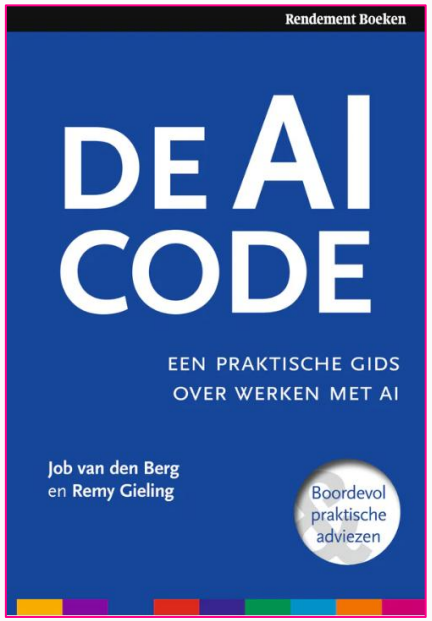
Verdiep je in AI – en experimenteer ermee!

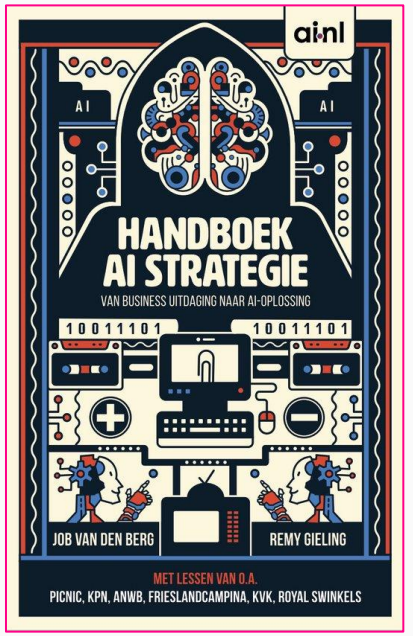
De vraag naar AI groeit ongekend, en dat geldt ook voor de ongelijkheidskloof die het creëert

Foto Wendy Barrows

Sander van 't Noordende topman uitzender Randstad waarschuwt voor een groeiende kenniskloof tussen werknemers die wel en niet geschoold zijn in kunstmatige intelligentie

Bron: De Volkskrant, 13 november 2024





Uit “Slimmer werken met AI”

“Studenten staken niet langer hun hand op, want waarom zou je je tijdens een college voor schut zetten als je het later aan de AI kon vragen?”

“De ‘transformer’ gebruikt een aandachtsmechanisme, waardoor de AI zich kan concentreren op de relevantste delen van een tekst. Het weegt het belang van verschillende frasen of delen van een tekstblok.”

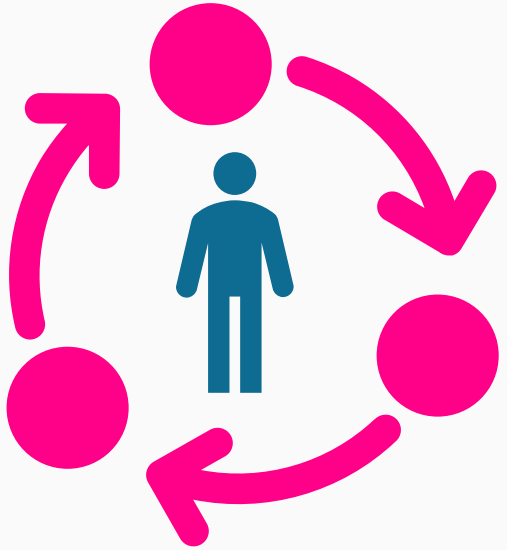
“Generatieve AI heeft in de basis geen bewustzijn en ook geen morele grenzen. AI wordt gefinetuned door Reinforcement Training from Human Feedback (RLHF)”

“Generatieve AI maakt een statistische berekening van het meest waarschijnlijke volgende woord of woorddeel in een tekst.”

Uit “Slimmer werken met AI”

Vier principes voor co-intelligentie

1. Nodig AI altijd uit aan tafel
2. Blijf the-human-in-the-loop
3. Behandel AI als een mens (maar zeg haar wat voor iemand ze is)
4. Ga ervan uit dat dit de slechtste AI is die je ooit zult gebruiken



Scan niet zomaar elke QR: Vraag je af of je de afzender kunt vertrouwen...

EnergyVision waarschuwt voor phishingmethode met QR-codes op laadpalen

Het Belgische energiebedrijf EnergyVision waarschuwt voor een vorm van phishing waarbij criminelen eigen QR-codes op laadpalen plakken, waarmee gebruikers zouden kunnen betalen voor het opladen van de auto. In werkelijkheid wordt hen geld afhandig gemaakt.

Bestuurders van elektrische auto's in Brussel kunnen op twee manieren betalen voor het opladen van hun auto aan een laadpaal. Dit kan met een laadpas, maar ook door een QR-code op de paal te scannen. Oplichters plakken echter op sommige palen een eigen, neppe QR-code over de echte heen, meldt VRT. Wie deze QR-code scant, komt op een betaalsite van de criminelen. EnergyVision noemt deze phishingmethode '*quishing*'.

Bron: Tweakers, 9 mei 2024

Vragen?

Dank, veel succes en wijsheid gewenst!

Nederlands compliance Instituut

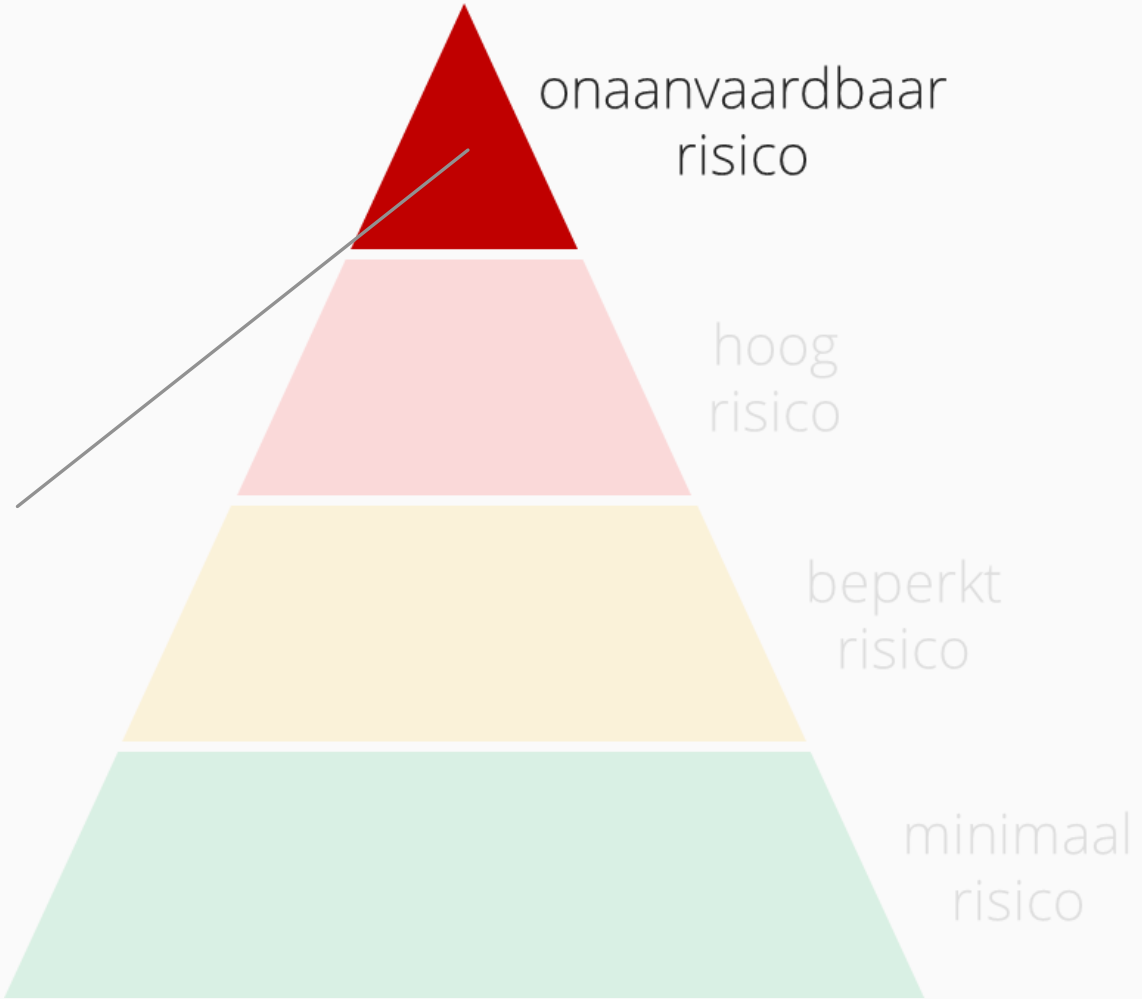
BIJLAGEN

Bron: Autoriteit Persoonsgegevens

Toepassingen met onacceptabele risico's zijn verboden

Onaanvaardbaar risico
Systemen die een onaanvaardbaar risico met zich meebrengen worden verboden. U mag deze systemen dus **niet aanbieden of gebruiken**. Dit verbod **gaat 2 februari 2025** in. Hieronder vallen bijvoorbeeld systemen en toepassingen die de vrije keuze van mensen te veel beperken, die manipuleren of die discrimineren.
Denk aan systemen die worden gebruikt in of voor:

- **'social scoring'** op basis van bepaald sociaal gedrag of persoonlijke kenmerken;
- **'predictive policing'** om strafbare feiten bij mensen te beoordelen of te voorspellen.
- het creëren of (via **scraping**) uitbreiden van databanken voor gezichtsherkenning;
- **manipuleren of misleiden van mensen;**
- **emotie-herkenning** in de werkomgeving en het onderwijs;
- **biometrische identificatie op afstand voor rechtshandhaving**. Hier zijn wel enkele uitzonderingen van toepassing;
- **biometrische categorisering**, waarbij mensen op basis van biometrische gegevens in bepaalde gevoelige categorieën worden ingedeeld.

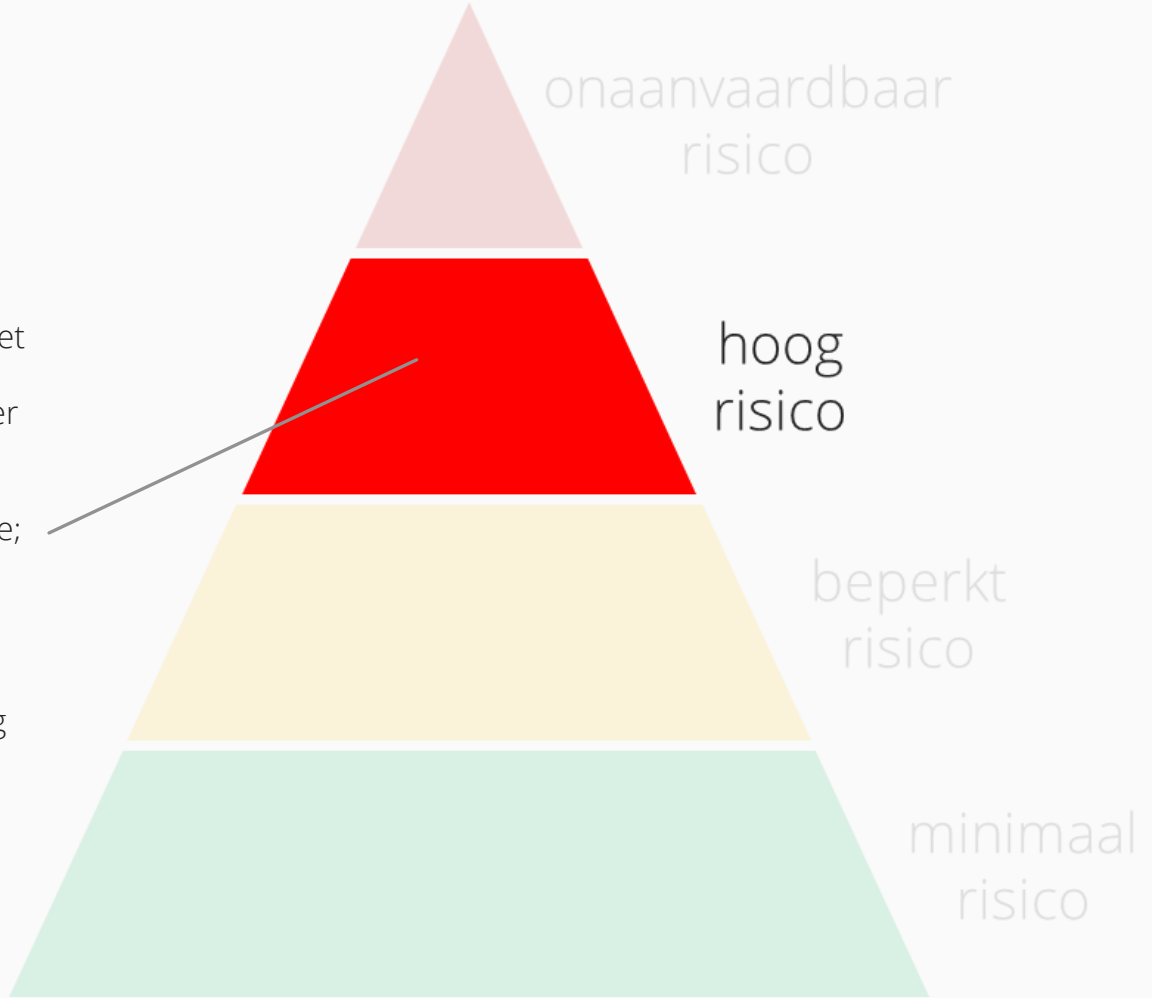


Bron: Autoriteit Persoonsgegevens

Hoog risico: Risicobeheersing verplicht

Hoog risico
Voor AI-systemen die vallen in de risicogroep 'hoog risico' gelden strikte verplichtingen. Wanneer u daar niet aan kunt voldoen, mag u het systeem niet in de handel brengen of gebruiken. Denk aan systemen die worden gebruikt in of voor:

- onderwijs of beroepsopleiding, waarbij het systeem de toegang tot het onderwijs en de loop van iemands beroepsleven kan bepalen. Bijvoorbeeld het beoordelen van examens;
- werkgelegenheid, personeelsbeheer en toegang tot zelfstandige arbeid, onder andere vanwege de aanzienlijke risico's voor toekomstige carrièrekansen en het levensonderhoud van mensen. Bijvoorbeeld een AI-systeem dat automatisch cv's selecteert voor de volgende ronde in een wervingsprocedure;
- essentiële particuliere en openbare diensten. Zulke systemen kunnen grote effecten hebben op bijvoorbeeld het levensonderhoud van mensen. Bijvoorbeeld software die bepaalt wanneer iemand wel of geen lening krijgt;
- rechtshandhaving, omdat dit afbreuk kan doen aan de grondrechten van mensen. Bijvoorbeeld een AI-systeem dat wordt gebruikt voor de beoordeling van de betrouwbaarheid van bewijsmateriaal;
- migratie, asiel en grenzen. Denk aan geautomatiseerde behandeling van asielaanvragen;
- rechtsbedeling en democratische processen, omdat het gebruik hiervan risico's kent voor de democratie en de rechtsstaat. Bijvoorbeeld een AI-systeem dat ondersteunt bij rechterlijke uitspraken.

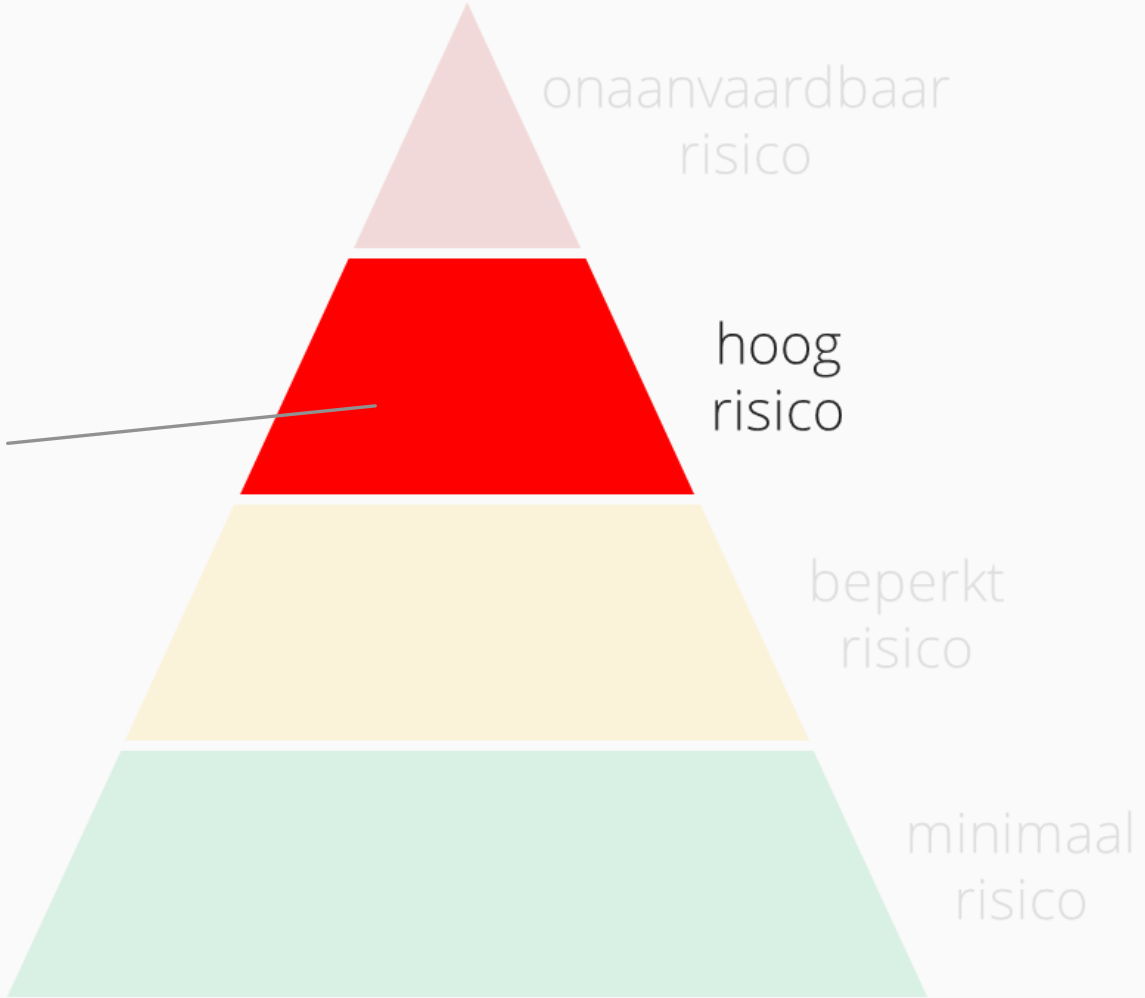


Bron: Autoriteit Persoonsgegevens

Hoog risico: Risicobeheersing verplicht

Hoog risico
Valt het systeem dat u wilt aanbieden of inzetten in de categorie 'hoog risico?' Dan gelden er regels om deze risico's te voorkomen of te beperken tot een aanvaardbaar niveau.
Bijvoorbeeld regels over risicobeheer, de kwaliteit van de gebruikte data, technische documentatie en registratie, transparantie en menselijk toezicht.

Overheden of uitvoerders van publieke taken moeten bij hoogrisicosystemen een 'grondrechteneffectbeoordeling' doen.



Bron: Tekst AI Verordening

Duidelijk?

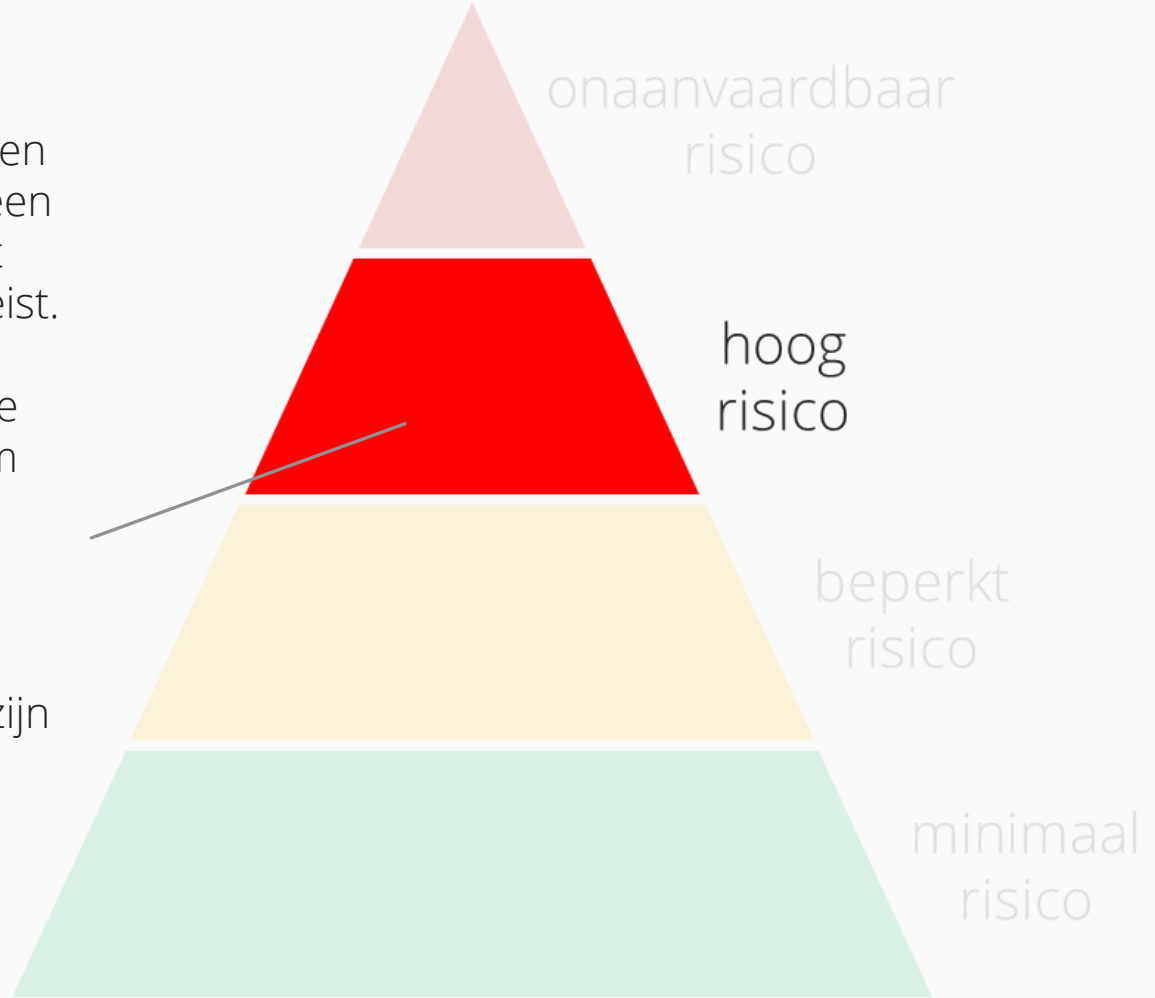
“De (...) bedoelde risicobeheers-maatregelen zijn zodanig dat het desbetreffende restrisico in verband met elk gevaar en het totale restrisico van de AI-systemen met een hoog risico aanvaardbaar wordt geacht.”

Bron: Tekst AI Verordening

Hoog risico: Risicobeheersing verplicht

Onder het **systeem voor risicobeheer** wordt verstaan een tijdens de gehele levensduur van een AI-systeem met een hoog risico doorlopend en gepland iteratief proces, dat periodieke systematische toetsing en actualisering vereist. Het omvat de volgende stappen:

1. Het vaststellen en analyseren van de bekende en de redelijkerwijs te voorziene **risico's** die het AI-systeem met een hoog risico kan inhouden voor de gezondheid, veiligheid of grondrechten (...)
2. Het **inschatten en evalueren** van de risico's (...)
3. Het evalueren van andere risico's die zich kunnen voordoen op basis van de analyse van de **data** die zijn verzameld door het systeem (...)
4. het vaststellen van gepaste en gerichte **risicobeheersmaatregelen** om (...) de vastgestelde risico's aan te pakken



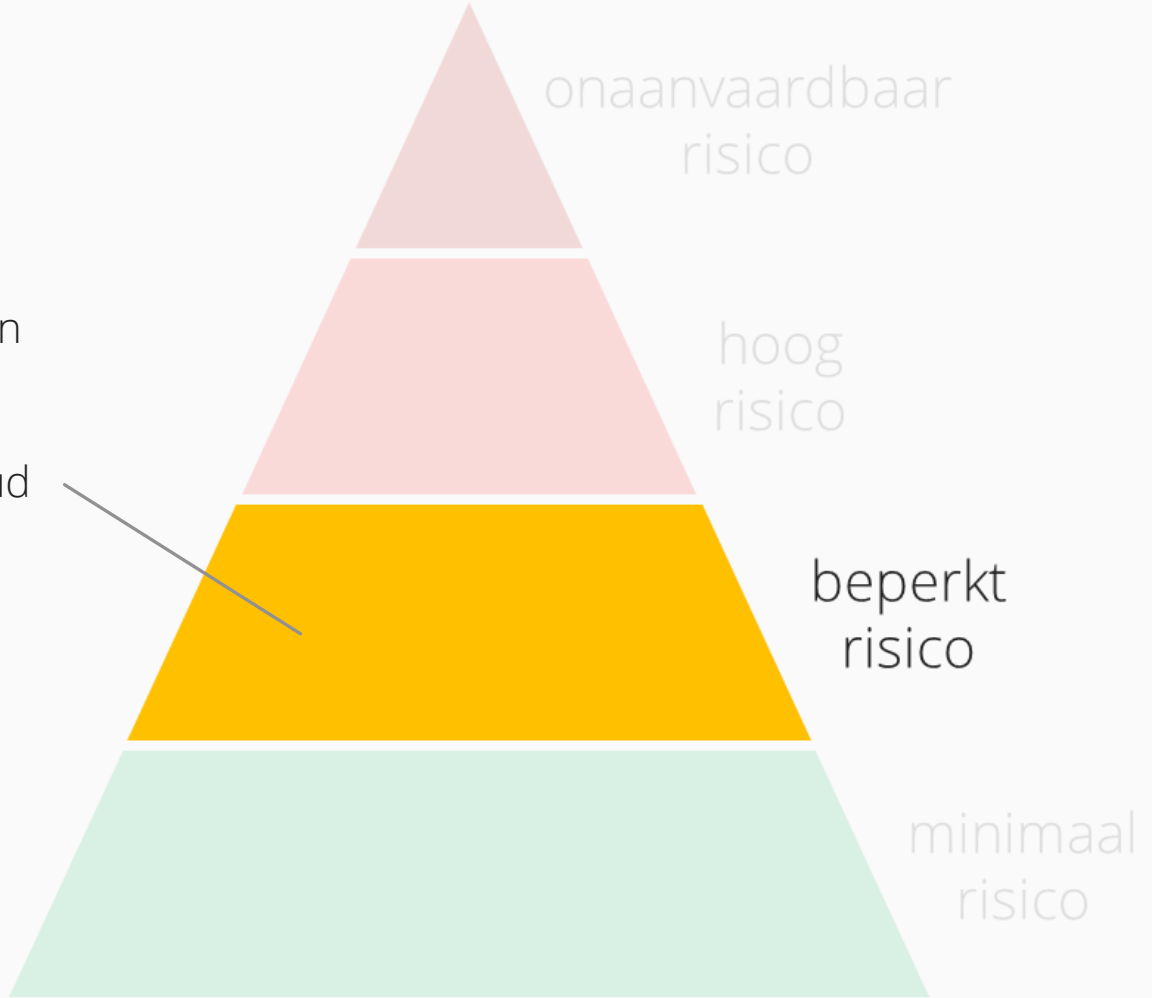
Bron: Autoriteit Persoonsgegevens

Beperkt risico: Transparantieverplichtingen

Beperkt risico

Voor AI-toepassingen met een beperkt risico gelden een aantal '**transparantieverplichtingen**'. Het gaat om AI-systemen die bedoeld zijn om met personen in contact te komen, zoals chatbots. En AI-systemen die zelf inhoud maken, zoals teksten en beeldmateriaal.

Als u deze systemen aanbiedt of gebruikt, dan moet u mensen erover **informeren** dat zij met AI te maken hebben.



Bron: Website EC

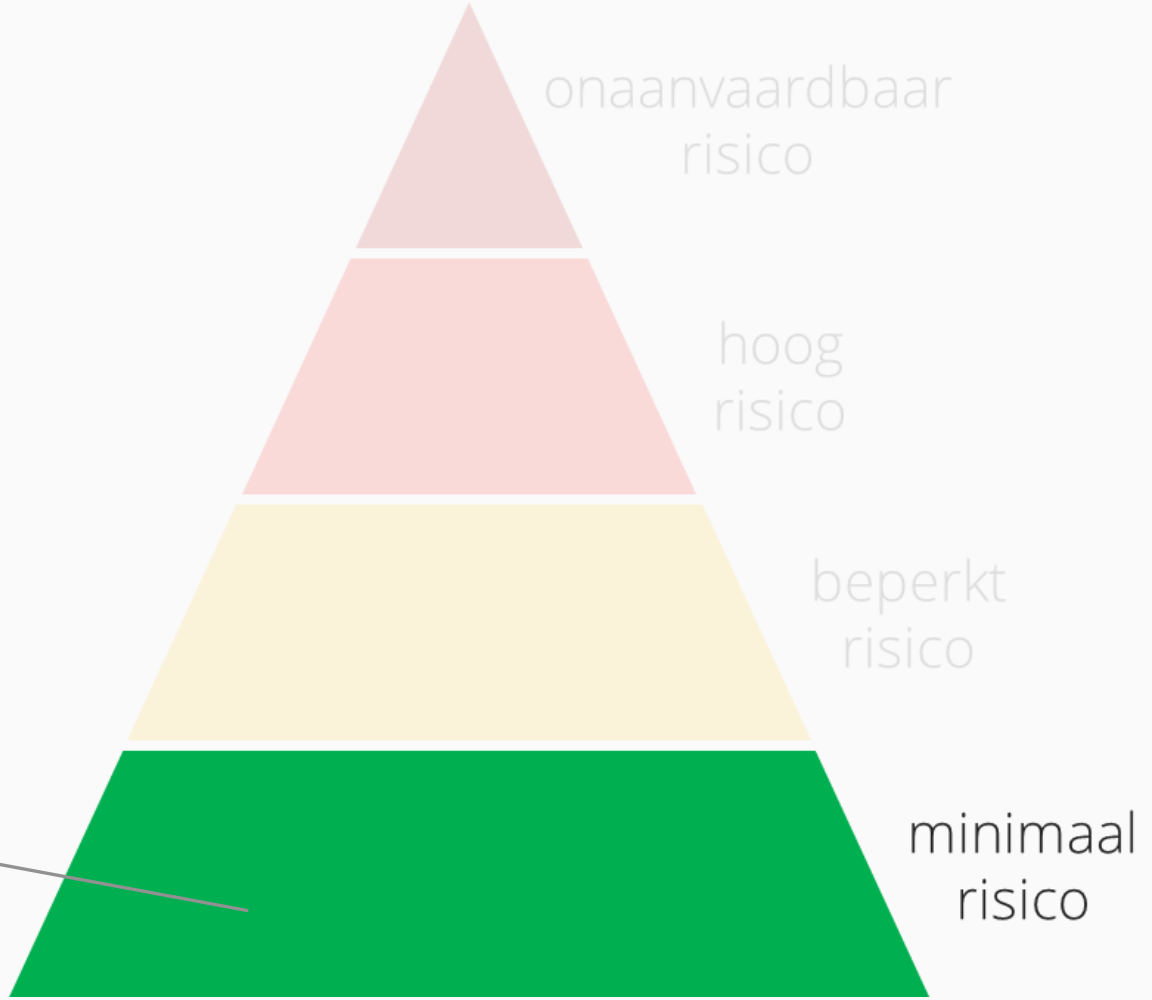
Minimaal risico: (Gratis) gebruik mogelijk

Minimaal risico

De AI-wet maakt het gratis gebruik van AI met een minimaal risico mogelijk.

Dit omvat toepassingen zoals AI-gestuurde videogames of spamfilters.

De overgrote **meerderheid van de AI-systemen** die momenteel in de EU worden gebruikt, valt onder deze categorie.



Een selectie van publicaties met waardevolle inzichten

Generative AI in a nutshell

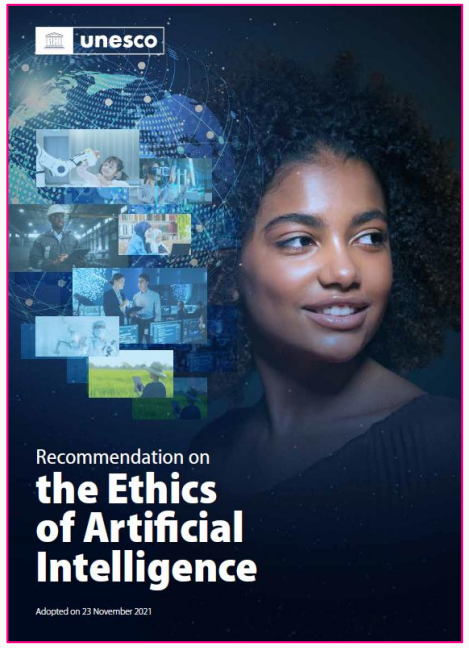
Transformers explained visually
(Inside an LLM)

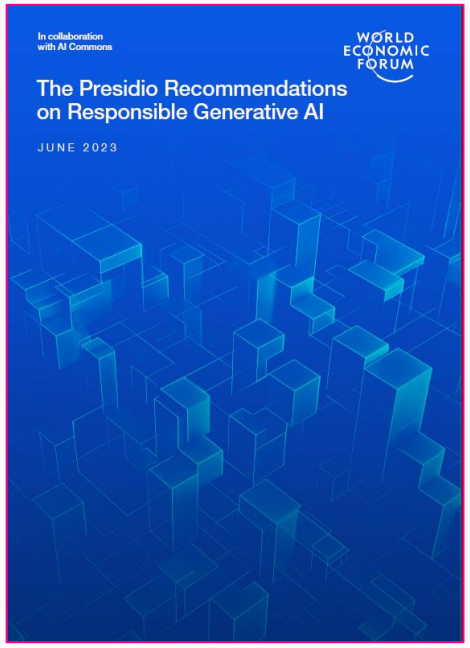
Een selectie van publicaties met waardevolle inzichten

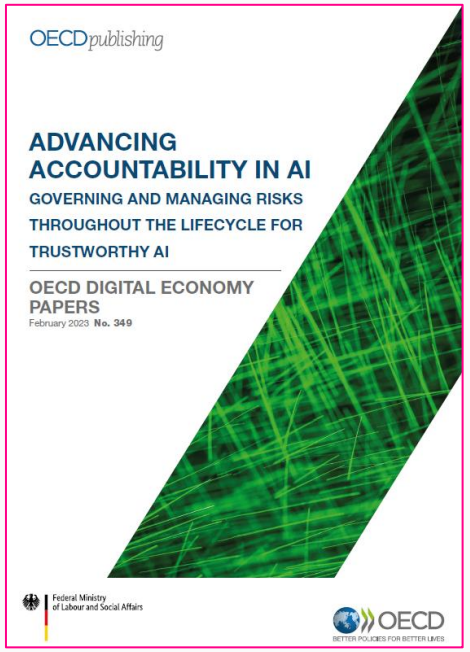
Stanford University USA: AI Trends 2024:

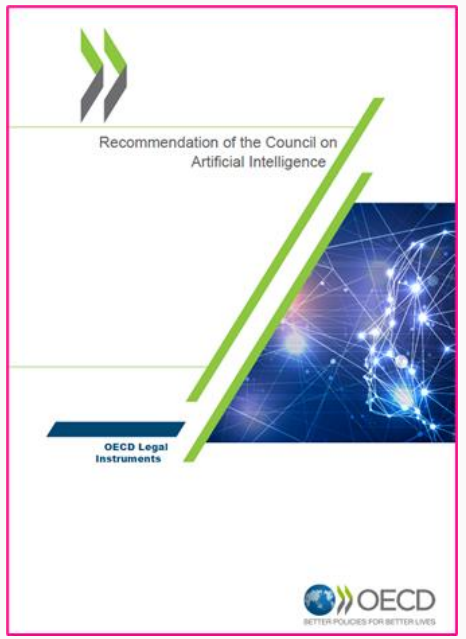
1. AI beats humans on some tasks, but not on all.
2. Industry continues to dominate frontier AI research.
3. Frontier models get way more expensive.
4. The United States leads China, the EU, and the U.K. as the leading source of top AI models.
5. Robust and standardized evaluations for LLM responsibility are seriously lacking.
6. Generative AI investment skyrockets.
7. The data is in: AI makes workers more productive and leads to higher quality work.
8. Scientific progress accelerates even further, thanks to AI.
9. The number of AI regulations in the United States sharply increases.
10. People across the globe are more cognizant of AI's potential impact—and more nervous.

Een selectie van publicaties met waardevolle inzichten

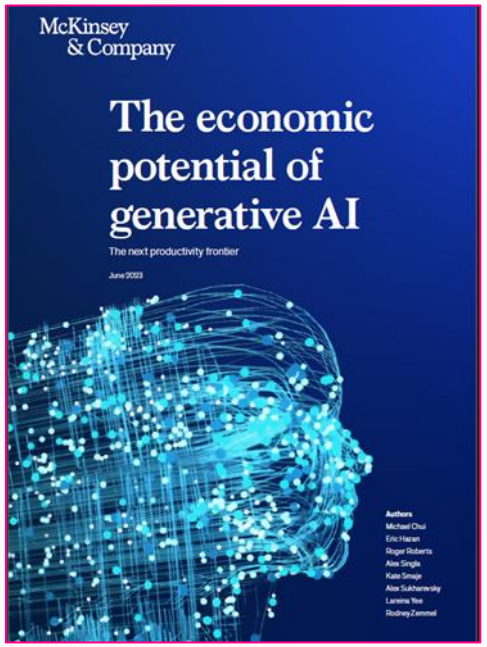


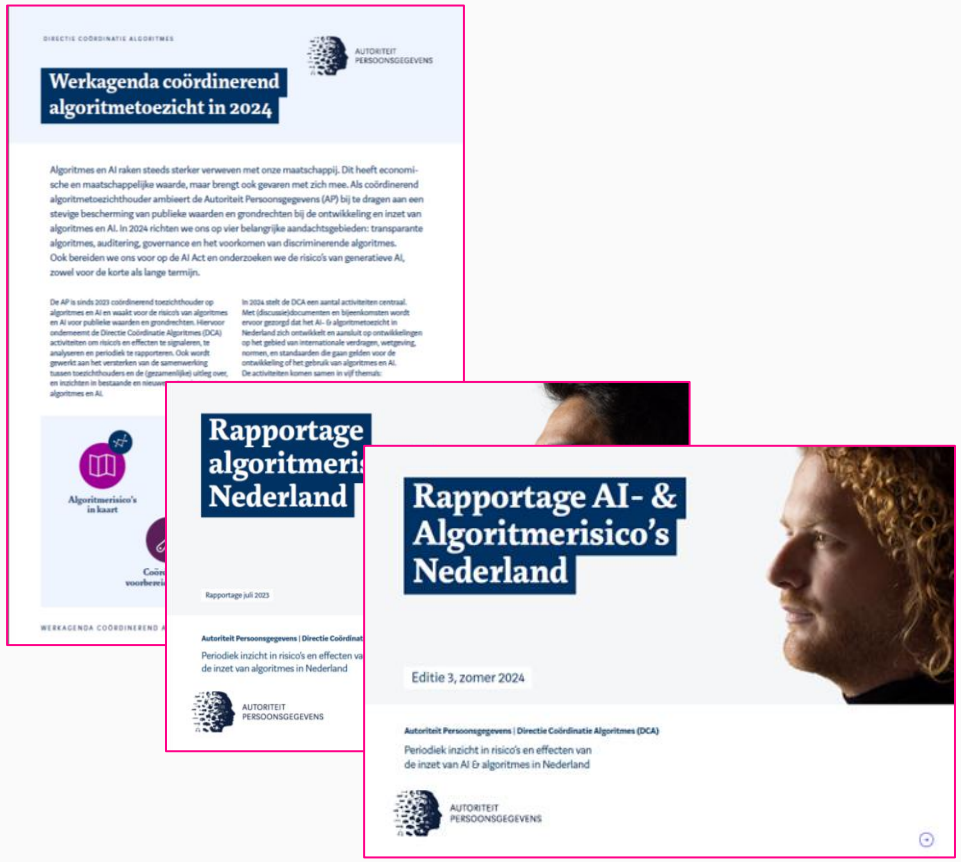




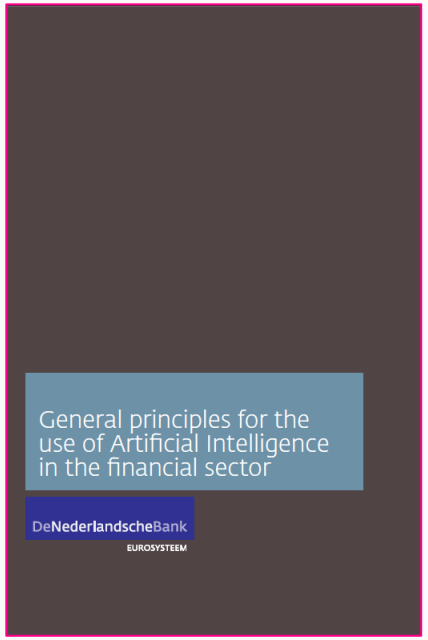


Een selectie van publicaties met waardevolle inzichten

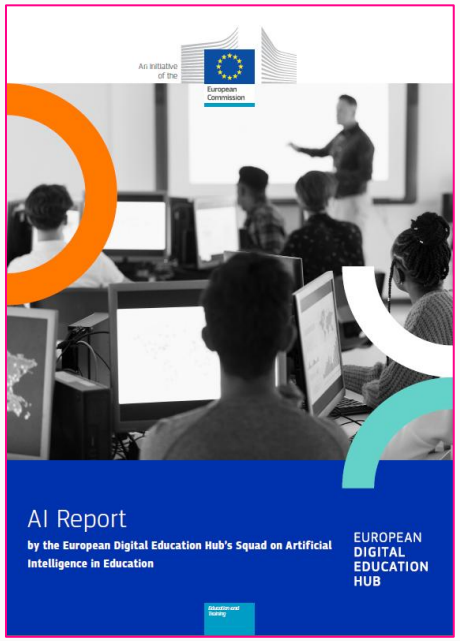


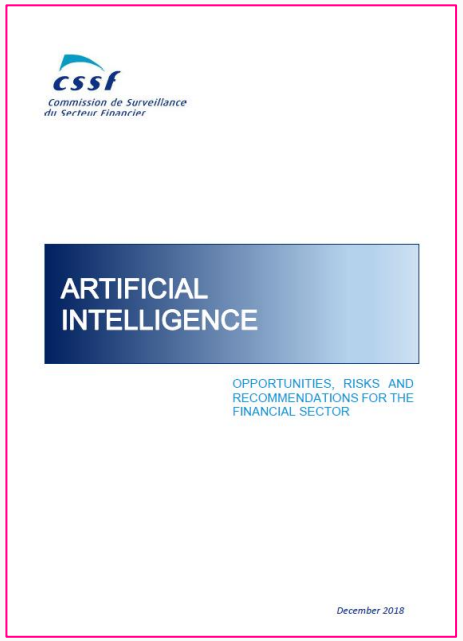


Een selectie van publicaties met waardevolle inzichten



Een selectie van publicaties met waardevolle inzichten





OECD Council (OESO) May 2024 Recommendation of the Council on AI

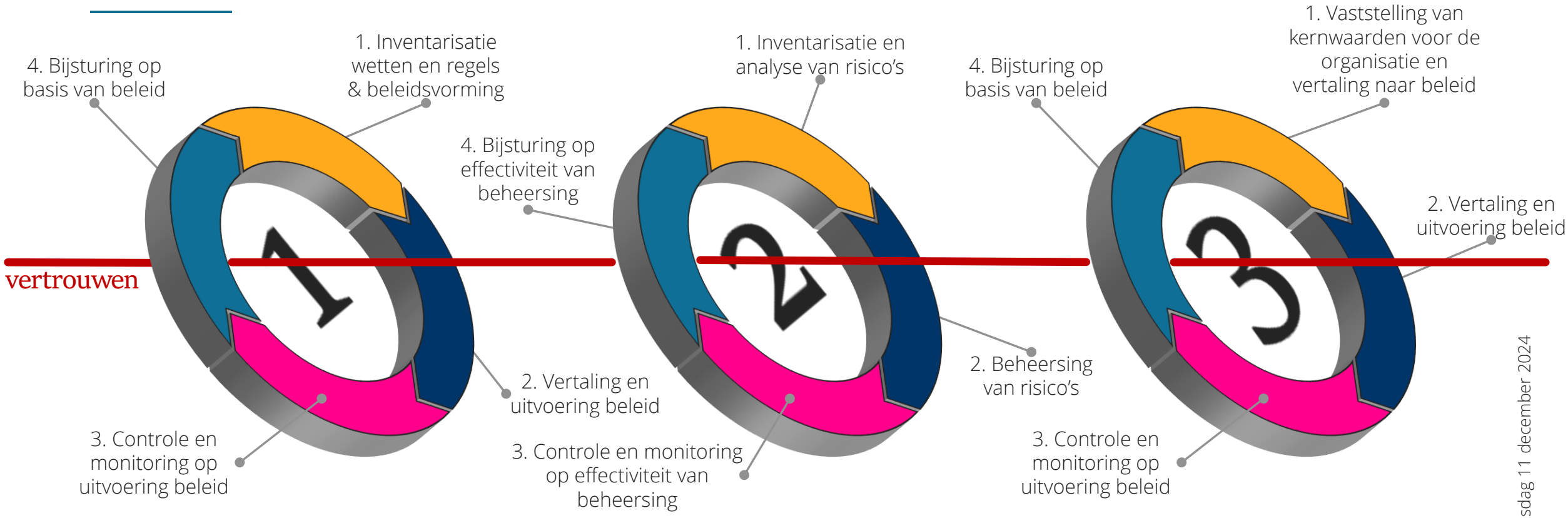
1. RECOMMENDS that Members and non-Members (…) adhering to this Recommendation promote and implement the following principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders.
   - Inclusive growth, sustainable development and well-being
   - Respect for the rule of law, human rights and democratic values, including fairness and privacy
   - Transparency and explainability
   - Robustness, security and safety
   - Accountability

OECD Council (OESO) May 2024 Recommendation of the Council on AI

2. RECOMMENDS that Adherents implement the following recommendations, consistent with the principles in section 1, in their national policies and international co-operation, **with special attention to small and medium-sized enterprises (SMEs).**
   - Investing in AI research and development
   - Fostering an inclusive AI-enabling ecosystem
   - **Shaping an enabling interoperable governance and policy environment for AI**
   - Building human capacity and preparing for labour market transformation
   - International co-operation for **trustworthy AI**

Bron: Nederlands Compliance Instituut

Het Twistermodel®: Groeien in compliance en integriteit



Transactionele aanpak
COMPLIANCE

Risicogedreven aanpak
INTEGRITEIT

Waardegedreven aanpak
ETHIEK