Zo werkt risicomanagement voor AI

Introductie

Articial Intelligence (AI) belooft de wereld te veranderen. De vele mogelijkheden van AI-toepassingen creëren nieuwe kansen, maar de razendsnelle ontwikkelingen roepen ook vragen en zorgen op. Bijvoorbeeld over privacy, manipulatie en discriminatie. In dit whitepaper deelt L2P inzichten en tips om als organisatie de voordelen van AI te benutten én de risico's te beheersen.

L2P heeft dit whitepaper speciaal ontwikkeld voor privacy officers, Functionarissen voor Gegevensbescherming, information security officers, Chief Information Security Officers, security officers, ICT-managers, security managers, risk managers, compliance officers en andere professionals die in het dagelijkse werk te maken hebben met risicomanagement rond AI.

Je krijgt een compact overzicht van de belangrijkste risico's en aandachtspunten voor compliance bij de toepassing van AI in je organisatie. Daarbij is speciale aandacht voor de vereisten van de AI-verordening. Daarnaast zetten we de zes thema's op een rij om effectief aan de slag te gaan met de succesvolle en verantwoorde inzet van AI.

Inhoudsopgave

L2P is een fullservice adviesbureau, gespecialiseerd in privacy, informatiebeveiliging en compliance. We vertalen de complexe vraagstukken van jouw organisatie naar praktische oplossingen waarmee je verder komt. Nu en in de toekomst.
De experts van L2P zorgen dat vraagstukken rond privacy, informatiebeveiliging en compliance goed zijn geregeld. Met onze jarenlange ervaring in de publieke en private sector snappen we de uitdagingen van jouw organisatie.
Heeft jouw organisatie privacy en informatiebeveiliging echt op orde? Onze adviseurs staan voor je klaar.

1 Risicomanagement voor AI

“Kort nadat de eerste auto's op de weg waren, vond het eerste auto-ongeluk plaats. Maar we verboden auto’s niet - we voerden snelheidslimieten, veiligheidsnormen, rijbewijzen, wetten tegen rijden onder invloed en andere verkeersregels in,” schreef Bill Gates in 2023 over de risico's van Artificial Intelligence (AI). De onderliggende gedachte is duidelijk: AI gaat niet meer weg en de opgave is nu om de risico's op de juiste manier te adresseren, zodat we gebruikmaken van de vele voordelen.

Gates voorziet verschillende acties om de adoptie van AI in goede banen te leiden. Hij pleit voor onder meer wet- en regelgeving, een open dialoog tussen politici en burgers, en de ontwikkeling van privacyvriendelijke AI-systemen en algoritmes zonder bias. Maar wat staat je als organisatie te doen om de risico's van AI beheersbaar te maken?

1.1 Risicomanagement: de basis

Risicomanagement en compliance zijn de basis voor de succesvolle en verantwoorde inzet van AI in je organisatie. Risicomanagement voor AI komt op hoofdlijnen overeen met de aanpak van andere ICT-toepassingen die (persoons)gegevens verwerken.

Het gaat erom dat je in ieder geval de volgende acties onderneemt:

- Inventariseer welke soorten informatie er worden verwerkt.
- Registreer wie er toegang hebben tot de informatie.
- Verken welke wet- en regelgeving van toepassing is.
- Bepaal wat er gebeurt als er iets misgaat.
- Neem maatregelen om de negatieve impact van de grootste risico's tegen te gaan.

1.2 Risicomanagement: de specifieke uitdagingen van AI

Om grip te krijgen in de risico's van AI is geen diepgaande kennis over de technische details noodzakelijk. Het is wel belangrijk om op de hoogte te zijn van een aantal AI-specifieke risico's:

- AI-modellen verwerken voortdurend nieuwe informatie. Er wordt bijvoorbeeld input van eindgebruikers verzameld om een model verder te trainen. De kwaliteit van de output van een applicatie of tool is dan ook mede afhankelijk van de juistheid van de input ("garbage in, garbage out").

- AI-chatbots worden steeds slimmer, maar momenteel zijn zogenoemde hallucinaties nog een reëel risico. Hierbij construeert AI zelf een antwoord op basis van associaties afkomstig uit het trainingsmodel. De werkwijze kan leiden tot onjuiste en zelfs onzinnige resultaten.

- AI-oplossingen worden ontwikkeld door de modellen te voorzien van zeer grote hoeveelheden (online) data. Daarbij wordt soms onvoldoende rekening gehouden met auteursrechten.

- Gebruikers van AI-systemen zijn zich niet altijd bewust van de risico's die ontstaan als persoonsgegevens en/of vertrouwelijke data worden ingevoerd in een chatbot. Bijvoorbeeld: het is mogelijk dat een medewerker vertrouwelijke gespreksverslagen in ChatGPT plaatst om een samenvatting te produceren, waardoor gevoelige gegevens in het bezit komen van OpenAI, het bedrijf achter ChatGPT.

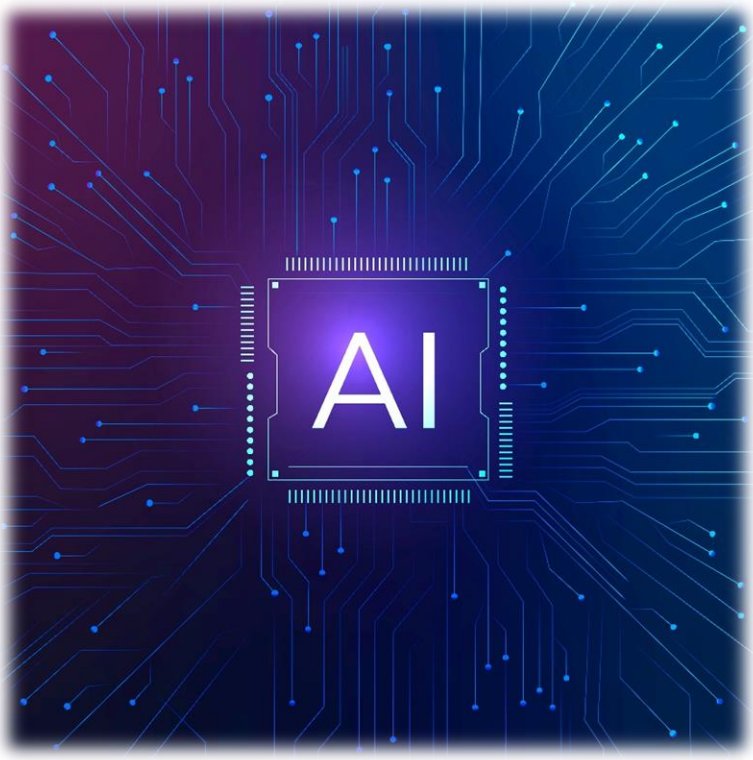
- De wereld van AI ontwikkelt zich razendsnel. Het is mogelijk dat een AI-product al kort na de aanschaf is veranderd, waardoor na de implementatie onvoorziene risico’s ontstaan, zoals onvoorspelbaar gedrag, moeilijk te stoppen processen, of besluiten en fouten die ondoorzichtig zijn.

- Steeds meer landen zijn wet- en regelgeving op het gebied van AI aan het ontwikkelen. Niet-naleving van de regels kan leiden tot hoge boetes en reputatieschade voor je organisatie.

1.3 Basisbeginselen voor risicomanagement rond AI

Uit de bovengenoemde AI-risico's zijn een aantal basale aandachtspunten voor het risicomanagement af te leiden:

- Zorg dat je vooraf weet welke informatie wordt gebruikt voor een AI-toepassing, wat ermee gebeurt en wat er vervolgens uit komt.
- Let erop dat de data waarmee een AI-toepassing is getraind representatief is en rechtmatig is verkregen.
- Maak de werkvloer steeds bewust van het juiste gebruik van AI-toepassingen.
- Blijf op de hoogte van eventuele veranderingen in AI-producten en -diensten, zodat de mogelijke implicaties van wijzigingen tijdig in beeld komen.
- Zorg dat het gebruik van AI-systemen onderhevig is aan een continue evaluatie en risicobeoordeling.



2 De naleving van de AI vordering

Een van de belangrijkste ontwikkelingen in het risicomanagement rond AI is de Europese AI-verordening, die eisen stelt aan de ontwikkeling en het gebruik van risicovolle AI-systemen. De scope van de wet betekent dat veel organisaties in de Europese Unie met nieuwe regels te maken krijgen. De wet is op 1 augustus 2024 in werking getreden, waarbij de verschillende onderdelen gefaseerd van kracht worden (zie bijlage).

Er is veel aandacht voor de vereisten van de AI-verordening. Voor AI-toepassingen die persoonsgegevens verwerken gelden ook de regels van de Algemene verordening gegevensbescherming (AVG). Dit betekent onder andere dat je een DPIA moet opstellen voordat AI-systemen in gebruik worden genomen.

2.1 De belangrijkste begrippen en regels van de AI-verordening

De AI-verordening geeft in artikel 1 de volgende definitie van een AI-systeem: "een op een machine gebaseerd systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat na het inzetten ervan 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 die van invloed kunnen zijn op fysieke of virtuele omgevingen."

Door de brede definitie van AI-systemen in de AI-verordening is het waarschijnlijk dat veel geavanceerde IT-systemen die beslissingen, voorspellingen, inhoud en aanbevelingen genereren onder de reikwijdte van de wet vallen. Daarom is het belangrijk om als organisatie het gebruik van AI nauwkeurig in beeld te brengen, en om te beoordelen welke acties nodig zijn om te voldoen aan de AI-verordening.

De AI-verordening is van toepassing op verschillende partijen in de Europese commerciële waardeketen, zoals:

- aanbieders: organisaties die AI-systemen (laten) ontwikkelen en op de markt brengen.
- importeurs en distributeurs van AI-systemen.
- de zogenoemde gebruiksverantwoordelijken: de organisaties die AI-systemen gebruiken.

De AI-verordening geldt ook voor aanbieders en gebruiksverantwoordelijken buiten de Europese Unie als de output in een Europese lidstaat wordt ingezet.

De AI-verordening regelt dat elke Europese lidstaat toezichthouders en marktautoriteiten aanwijst die de naleving van de wet handhaven, die bij overtredingen boetes kunnen opleggen en die organisaties kunnen verplichten om een product uit de handel te nemen.

- In Nederland is de Autoriteit Persoonsgegevens (AP) coördinerend toezichthouder op AI en algoritmes. Het College voor de Rechten van de Mens (CRM) is de toezichthouder op het gebied van de grondrechtenbescherming.
- Markttoezichthouders controleren of AI-systemen met een hoog risico voldoen aan de eisen van de AI-verordening en de normen van sector. In Nederland is het marktoezicht belegd bij de Nederlandse Voedsel- en Warenautoriteit (NVWA), de Inspectie Gezondheidszorg en Jeugd (IGJ), de Autoriteit Financiële Markten (AFM), De Nederlandsche Bank (DNB) en de Rijksdienst voor Infrastructuur (RDI).

2.2 Risicogerichte maatregelen

De AI-verordening is risicogericht: de regulering is bedoeld voor AI-systemen met een risico voor personen en de samenleving. Naarmate het risiconiveau van een AI-toepassing toeneemt, zijn er strengere regels.

De wet onderscheidt twee concrete risiconiveaus, een onaanvaardbaar en een hoog risico, en stelt verder verplichtingen aan systemen die rechtstreeks met personen communiceren of op een andere manier interageren.

- AI-systemen met een onaanvaardbaar risico zijn verboden. Het gaat onder andere om systemen die de vrije keuze van mensen te veel beperken, die mensen manipuleren, of die discriminerend zijn. Voorbeelden zijn het beoordelen van mensen op basis van hun sociale gedrag of persoonlijke kenmerken, het voorspellen van crimineel gedrag en emotieherkenning op de werkplek en in het onderwijs.. AI-systemen die een onaanvaardbaar risico met zich meebrengen zijn sinds februari 2025 verboden.

- Voor AI-systemen met een hoog risico gelden strikte regels. Aan de AI-verordening is een bijlage toegevoegd van acht toepassingsgebieden die onder deze categorie vallen. Denk aan AI die wordt ingezet in het onderwijs en recruitment. De verplichtingen voor AI met een hoog risico gelden vanaf augustus 2026.

- Dan zijn er nog “bepaalde AI-systemen”, ofwel systemen die voor directe interactie met personen zijn bedoeld. Deze systemen moeten voldoen aan een aantal transparantie-eisen. Toepassingen zoals chatbots moeten personen informeren over het feit dat ze met een AI en niet een mens praten. Tools die teksten en beelden moeten de geproduceerde content op de een of andere manier ‘watermerken’ zodat duidelijk is dat AI het gemaakt heeft. De transparantieverplichtingen gelden vanaf augustus 2026.

Verplichtingen voor AI-modellen voor algemene doeleinden.
De focus van de AI-verordening ligt op AI-systemen en toepassingen. Maar er zijn ook vereisten voor de zogenoemde AI-modellen voor algemene doeleinden. Het betreft modellen die teksten, audio, video en computercode genereren en daardoor veelzijdig inzetbaar zijn. De modellen zijn geen volwaardige tools, maar staan vaak aan de basis van AI-systemen. Hieronder vallen de grote generatieve modellen die gebruikt worden om ChatGPT en Copilot te laten draaien.

- Aanbieders van AI-modellen voor algemene doeleinden zijn op grond van de AI-verordening verplicht om technische documentatie op te stellen. Onder andere over het trainings- en testproces en de resultaten van evaluaties. Aanbieders moeten ook zorgen voor actuele informatie die de ontwikkelaars van AI-systemen inzicht geven in de (on)mogelijkheden van de modellen. Verder moeten zij verzekeren dat zij tijdens het trainen van hun modellen auteursrechten respecteren.

- Aan AI-modellen die systeemrisico's creëren, zoals de verspreiding van ongepaste en gevaarlijke content, stelt de wet aanvullende eisen. Deze categorie is in het leven geroepen om rekening te houden met AI-modellen die een grote impact kunnen hebben op de samenleving. Dergelijke modellen zijn buiten de hoog-risico categorie gehouden omdat ze strikt gezien nog geen zelfstandige AI-systemen zijn, maar wel veel gebruikt worden en een brede toepasbaarheid hebben. Met de systeemrisico-kanttekening probeert de AI-verordening eventuele gevolgen voor bijvoorbeeld de volksgezondheid, veiligheid of de democratie te beperken. Aanbieders zijn verplicht om onder andere regelmatig evaluaties uit te voeren om kwetsbaarheden te identificeren, om incidenten te documenteren, en om maatregelen te nemen voor cyberbeveiliging.

De risico's van AI die is verwerkt in onder meer spamfilters en inparkeersystemen in auto's worden gezien als minimaal. Voor deze toepassingen geldt de AI-verordening niet.

2.3 Specifieke maatregelen

De AI-verordening verplicht organisaties om een aantal concrete maatregelen te nemen.

Fundamental Rights Impact Assessment

Bij het gebruik van een AI-systeem met een hoog risico in de publieke sector is een Fundamental Rights Impact Assessment (FRIA) verplicht. De effectbeoordeling geeft inzicht in de mogelijke gevolgen van het gebruik van AI voor de grondrechten van burgers, waaronder privacy en veiligheid. Een FRIA is verplicht voor overheidsinstanties en voor private partijen die publieke diensten aanbieden.

AMA en FRIA

In Nederland zijn organisaties sinds 2022 verplicht om een Impact Assessment voor Mensenrechten bij de inzet van Algoritmes (IAMA) uit te voeren als algoritmes worden ingezet om evaluaties van en/of beslissingen over mensen te maken. Een IAMA is gericht op een analyse van algoritmes. Een algoritme is niet per definitie een AI-systeem of -toepassing. Om te voldoen aan alle verplichtingen van artikel 27 van de AI-verordening is het daarom noodzakelijk om een IAMA aan te passen, of om alsnog een FRIA uit te voeren.

Conformiteitsbeoordeling

Voor AI-systemen met een hoog risico is een conformiteitsbeoordeling nodig. Zo'n risicoanalyse moet zijn afgerond voordat een product of dienst op de markt verschijnt of in gebruik wordt genomen. Artikel 43 van de AI-verordening benoemt de eisen waaraan een conformiteitsbeoordeling moet voldoen. Een organisatie kan ervoor kiezen om de beoordeling zelf uit te voeren, of hiervoor een externe partij in te schakelen.

Let op:

- Net als een DPIA is een conformiteitsbeoordeling geen eenmalig proces. Als het AI-systeem significante wijzigingen ondergaat, is er opnieuw een beoordeling nodig.

- Als een AI-systeem met een hoog risico persoonsgegevens verwerkt, dan is naast een conformiteitsbeoordeling ook een DPIA verplicht.

Transparantie-eisen

Aan AI-systemen met een beperkt risico stelt de AI-verordening speciale transparantie-eisen. De transparantieverplichtingen gelden voor aanbieders én gebruikers. Bijvoorbeeld voor organisaties die gebruik maken van AI-chatbots en/of door AI gegenereerde (of gemanipuleerde) teksten en beelden. De transparantieverplichting betekent dat je mensen op de hoogte stelt dat ze te maken hebben met een AI-systeem en/of met AI-gegenereerde content. Dit kan bijvoorbeeld in een bijschrift bij een afbeelding aan te geven dat het beeld door AI is gemaakt.

AI-geletterdheid

Alle organisaties die AI-systemen ontwikkelen of gebruiken, moeten sinds februari 2025 zorgen dat medewerkers "AI-geletterd" zijn. Medewerkers moeten over voldoende kennis, vaardigheden en begrip beschikken om AI op een verantwoorde manier in te zetten. Daarbij is aandacht nodig voor de technische werking van AI, maar ook van praktische en ethische aspecten.

3. Aan de slag: zes thema's

De verantwoorde omgang met AI is geen eenvoudige opgave. Bij het risicomanagement van AI-toepassingen spelen allerlei praktische, technische en juridische vraagstukken.

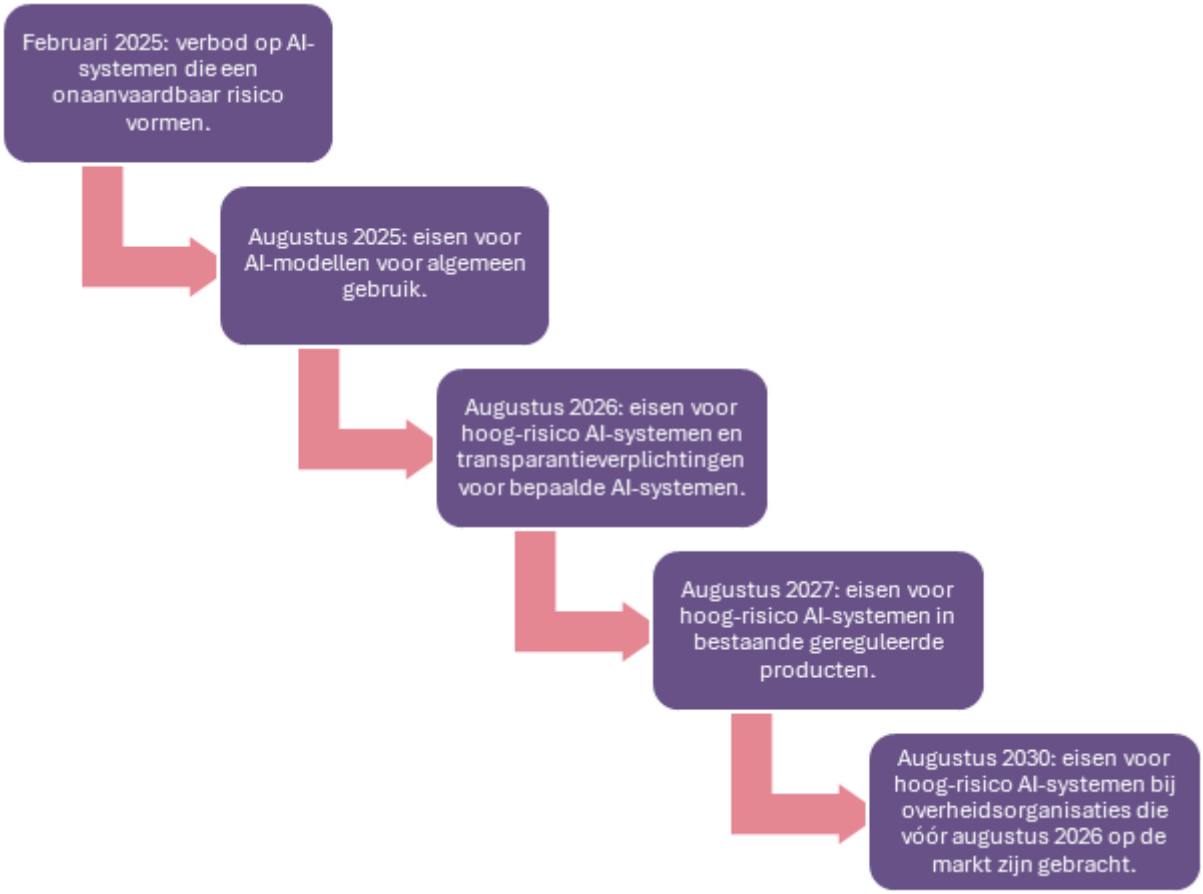
Voor het risicomanagement rond AI is aandacht nodig voor de volgende zes thema's:

- **GOVERNANCE.** Ontwikkel een AI-strategie over de doelen, uitgangspunten en verantwoordelijkheden. Organiseer de verantwoordelijkheden voor de uitvoering ervan. Stel duidelijke richtlijnen en procedures op, bijvoorbeeld met het oog op rapportages aan stakeholders en toezichthouders over incidenten en de afhandeling daarvan.
- **CLASSIFICATIE**. Breng in kaart welke AI-systemen er binnen je organisatie aanwezig zijn en hoe deze worden gebruikt. Bepaal welke systemen binnen de reikwijdte van de AI-verordening en AVG vallen en expliciteer welke wettelijke vereisten daarvoor gelden.
- **LEVERANCIERSMANAGEMENT**. Inventariseer op welke manier je leveranciers gebruik maken van AI-modellen en -systemen. Check of aanbieders voldoen aan de wettelijke verplichtingen.
- **RISICO-ANALYSE EN -MITIGATIE.** Breng de risico's van AI-toepassingen in je organisatie in kaart. Voer - waar relevant of nodig - DPIA's en/of FRIA's uit. Ontwikkel en implementeer de maatregelen om de geïdentificeerde risico’s te verminderen.
- **MONITORING.** Voer regelmatig evaluaties, audits en/of DPIA's uit. Monitor het correcte gebruik van AI-systemen, en analyseer en evalueer eventuele incidenten en storingen. Controleer regelmatig of instructies nog actueel zijn. Beschrijf daarbij ook de processen om AI-gebruikers en andere stakeholders te informeren over updates.
- **OPLEIDING EN TRAINING**. Investeer in opleidingen, trainingen en campagnes die zorgen dat medewerkers over voldoende AI-geletterdheid en risico-awareness beschikken.

Wil je weten hoe L2P jouw organisatie kan helpen met een of meerdere van deze 6 thema's? Neem dan vrijblijvend contact met ons op.

Bijlage 1

De vijf fasen van de inwerkingtreding van de AI-verordening



Bijlage II

Drie categorieën van AI

AI is in de mode en de *buzz words* vliegen ons om de oren. We geven een kort overzicht van drie hoofdcategorieën van AI en de betekenis ervan.

NARROW AI, *ook wel weak AI of Artifical Narrow Intelligence (ANI) genoemd*
ANI is ontwikkeld voor een specifieke taak. Denk aan chatbots zoals ChatGPT, generatieve systemen zoals Bard, Llama en MidJourney, en Microsoft Copilot. Ook AI-systemen die spraak naar tekst omzetten, tools voor beeldherkenning navigatie-applicaties zoals Google Maps zijn narrow AI.

In narrow AI worden verschillende technieken gebruikt die zijn gebaseerd op machine learning.

- Neural networks imiteren de werking van het menselijk brein. Het oorspronkelijke Google Translate was gebaseerd op een neuraal netwerk.
- De laatste jaren zijn de large language models (LLM) in opmars. De modellen zijn getraind op zeer grote hoeveelheden tekstdata om menselijke taal te begrijpen en te genereren. Een voorbeeld van chatbots die gebruik maken van LLM is ChatGPT. Ook Llama, een product van Meta, maakt gebruik van LLM.
- Generatieve AI (GenAI) is een vorm van narrow AI die specifiek is gericht op het creëren van nieuwe content, zoals teksten, afbeeldingen, video's, audiobestanden en programmeercodes. Voorbeelden zijn DALL-E, Midjourney, Google Gemini en Microsoft Copilot.

GENERAL AI, *ook wel strong AI of Artificial General Intelligence (AGI) genoemd*
AGI is in staat om - bijna - menselijk te denken, te redeneren en te beslissen. Een AGI-toepassing beschikt over lerend vermogen en verbetert daarmee de eigen prestaties. AGI belooft een breed spectrum aan taken uit te voeren. Denk aan het voeren van een gesprek, het besturen van een auto en het schrijven van een roman. Momenteel zijn dergelijke denk- en redeneer-activiteiten - tot op een zekere hoogte - mogelijk door verschillende soorten ANI in te zetten. Eén systeem dat het allemaal kan, is nog toekomstmuziek.

SUPER AI, *ook wel Artificial Super Intelligence (ASI) genoemd*
ASI is de volgende stap na AGI. Waar AGI het menselijk vermogen imiteert, is ASI de overtreffende trap die het menselijk vermogen overstijgt. We zien ASI nu vooral in science fiction.

Colofon

Dit whitepaper is in 2025 opgesteld door experts van L2P.

L2P stelt het document uitsluitend beschikbaar om een overzicht te geven van een aantal stappen in de analyse van het volwassenheidsniveau van de informatiebeveiliging van een organisatie.
Gegevens in deze publicatie mogen niet als juridisch advies worden beschouwd.
Hoewel L2P de inhoud van het document met de grootst mogelijke zorgvuldigheid heeft opgesteld, garanderen we nooit dat de publicatie vrij is van onjuistheden of onvolledigheden. L2P aanvaardt geen enkele aansprakelijkheid voor schade op welke manier dan ook ontstaan door gebruik (in welke vorm dan ook), onvolledigheid of onjuistheid van dit document.

L2P
Stadsplateau 7
3521 AZ Utrecht

+31 (0) 26 848 3118
info@l2p.nl