Rapportage AI & Algoritmes Nederland

Zomer 2025 (Editie 5, juli 2025)

Periodiek inzicht in trends, risico's en beheersing van de inzet van AI & algoritmes in Nederland

AP | bescherming in een digitale wereld

Autoriteit Persoonsgegevens
Directie Coördinatie Algoritmes (DCA)

Inhoudsopgave

Kernboodschappen

1. De ontwikkeling en inzet van AI-systemen om emoties te herkennen groeit, terwijl de werking wordt betwijfeld en de inzet risicovol is.

Van steeds meer AI-systemen wordt geclaimd dat ze emotionele toestanden, zoals blijdschap, boosheid of stress, kunnen herkennen op basis van biometrie. Denk hierbij aan het herkennen van emoties voor inzicht in koopgedrag van consumenten, of voor inzicht in emoties van patiënten voor verbetering van de zorg. De systemen zijn gebouwd op omstreden aannames over emoties en indicatoren van emotionele toestanden en de meetbaarheid daarvan. Daarom is het twijfelachtig of ze ook echt emoties kunnen meten. Als de systemen toch worden ingezet, kan dit grondrechten en publieke waarden schenden. Denk aan de inperking van vrijheid en menselijke autonomie, discriminatie en privacy-schendingen. Lees in hoofdstuk 3 meer over AI en emotieherkenning.

2. De AP heeft toepassingen van emotie-herkenning voor klantenservices, in wearables en in taalmodellen onder de loep genomen; het brede palet aan toepassingen leidt in bepaalde gevallen tot risico's die aandacht vragen van ontwikkelaars, gebruikers, toezicht-houders en beleidsmakers.

Uit de verdieping blijkt bijvoorbeeld dat niet altijd duidelijk is hoe emoties en stress worden herkend of hoe effectief het is. Ondanks de groei van deze toepassingen, weten mensen toch niet altijd dat emotieherkenning wordt ingezet, of op basis van welke data. Ontwikkelaars en gebruikers moeten zich bewust zijn van de risico's die elke specifieke toepassing met zich meebrengt. Verschillende toepassingen vallen straks onder specifieke regelgeving. Maar het is een andere vraag of het *wenselijk* is om deze systemen in te zetten. In het onderwijs en op de werkplek is AI voor emotieherkenning sinds februari 2025 verboden, maar waar het toegestaan is onder specifieke voorwaarden zal het uiteindelijk een politieke afweging zijn om te bepalen in welke gevallen en mate de inzet van deze systemen wenselijk zijn. Lees in hoofdstuk 4 meer over emotieherkenning in de praktijk.

3. De AP roept organisaties op om zich kritisch te verhouden tot de inzet van toepassingen van emotieherkenning.

Organisaties moeten zich bewust zijn van de beperkingen van emotieherkenning en een bewuste afweging maken of de inzet ervan passend en proportioneel is. Als deze systemen worden ingezet moeten organisaties transparant zijn over het gebruik richting degene wiens emotie wordt geanalyseerd. Bovendien is toestemming van de personen die geanalyseerd worden een belangrijke eerste stap voor organisaties om de systemen verantwoord in te zetten. In hoofdstuk 4 worden drie praktijkvoorbeelden van emotieherkenning besproken. Maar ook voor andere toepassingen is werken aan een verantwoorde, bewuste en transparante inzet belangrijk.

4. In Nederland wordt met toewijding gewerkt aan betrouwbare AI, maar als samenleving moeten we een tandje bijschakelen om de kansen van AI op een verantwoorde manier te kunnen benutten.

Betrouwbare inzet van algoritmes en AI is mogelijk en vraagt om aandacht voor bescherming, veiligheid en transparantie. Dit betekent dat we als samenleving lessen moeten leren uit incidenten. Gezien de snelheid

en omvang waarmee Nederlandse organisaties AI en algoritmes adopteren lijkt het waarschijnlijk dat veel incidenten niet opgemerkt worden, niet gemeld worden en dat lessen daaruit ook niet gedeeld worden. Dat incidenten voorkomen is onvermijdelijk in deze fase van de maatschappelijke transitie. Omgaan met incidenten en het leren van de aard, oorzaak en schade is essentieel om organisaties te laten groeien in verantwoordelijke omgang. Toezichthouders kunnen samen met wetgevers en het toezichtsveld werken aan een passende omgeving waarbinnen deze kennis gedeeld kan worden zonder bedrijfsgeheimen te onthullen.

5. Het is inmiddels voor iedere organisatie mogelijk om, met behulp van het steeds rijkere pakket aan hulpmiddelen, doelgericht te werken aan AI-volwassenheid.

Van algoritmeregistratie, bias-toetsing en het inrichten van kwaliteitsmanagementsystemen tot het uitvoeren van grondrechtenbeoordelingen en het transparant maken van de inzet - om daar te komen moeten organisaties investeren in mensen en middelen en zijn meetbare en concrete organisatiedoelen nodig. Bijvoorbeeld een ambitie om jaarlijks de meest kritische algoritmische processen en AI-systemen te auditeren op belangrijkere criteria zoals organisatorische grip, robuustheid, bias en fairness, cyberveiligheid en willekeur. Dit is een uitdaging voor organisaties, waarbij een beroep wordt gedaan op de verantwoordelijkheid en eigen inzicht van een organisatie. Volwassen organisaties zijn transparant over de inzet, hebben robuuste systemen en processen voor beheersing, en nemen interne controle, externe controle en toezicht serieus.

6. Vormgeving van menselijke toetsing en het tegengaan van discriminatie blijft een belangrijke uitdaging.

In Nederland zijn we ons zeer bewust van het feit dat zowel personen als algoritmes kunnen discrimineren. Dit laat duidelijk het belang zien om zowel de menselijke als algoritmische bias in processen terug te dringen. De wisselwerking tussen mensen en algoritmes is hiervoor van belang en moet goed vormgegeven zijn om discriminatie en bias zoveel mogelijk te voorkomen. De Tweede Kamer heeft onlangs een motie aangenomen, gericht op het 'blind' beoordelen van burgers als uitgangspunt. Bijvoorbeeld bij fraudedetectie. Vanuit het perspectief van de AP als toezichthouder is het noodzakelijk om ook naar de besluitvormingsketen als geheel te kijken, aselecte steekproeven toe te voegen, de uitlegbaarheid en het kunnen aanvechten van de risicoselectie te waarborgen.

7. AI speelt een belangrijke rol in geopolitieke ontwikkelingen.

Succesvolle inzet op AI wordt gezien als essentieel voor economische ontwikkeling. Het bevorderen van innovatie en het AI-ecosysteem is in rap tempo hoog op de politieke agenda beland. Waar twijfels zijn over het beperken van innovatie door regels, kijkt men al snel naar deregulering. Maar om te beschermen tegen uitwassen wordt tegelijk ook nieuwe regelgeving ontwikkeld of geïmplementeerd. Ook verschilt het perspectief op de juiste balans tussen innovatie en bescherming per land en regio. Daarnaast wordt AI in groeiende mate onderdeel van nationale strategieën, variërend van de zorg tot defensie. Hierbij speelt de afhankelijkheid van grote spelers en de groeiende wens voor strategische autonomie een steeds grotere rol. Ook Europees wordt aandacht besteedt aan dit onderwerp, bijvoorbeeld middels de AI-strategie van de Europese Commissie met bijbehorende financieringsambities.

8. Het is zaak tempo te houden of te versnellen bij het ontwikkelen en implementeren van beleid voor AI.

Om de kansen van AI te benutten en de risico's te beheersen blijft een nationale overkoepelende AI-strategie noodzakelijk. Het is belangrijk om algoritmeregistratie voor (semi-)publieke organisaties te verplichten wanneer deze impactvolle algoritmes of AI inzetten. Ook beveelt de AP aan om periodieke audits te verplichten voor deze algoritmes en AI. Om verantwoorde innovatie in goede banen te leiden moeten er voldoende middelen voor AI-toezicht in Nederland zijn bij alle betrokken toezichthouders en inspecties. Om toezicht effectief gezamenlijk te organiseren en mee te laten bewegen met de snelle technologische ontwikkelingen, zijn investeringen in samenwerkingsverbanden onmisbaar. De AP adviseert om bij publieke investeringen in innovatie een vastgesteld percentage te reserveren voor interne controle, externe controle en toezicht. Ook bij private investeringen is het aan te bevelen om een deel te reserveren voor risicomanagement inclusief interne en externe controle. Het goed functioneren van deze lagen is essentieel voor versnelling van innovatie en inzet.

9. De AI-verordening wordt steeds concreter, waarbij standaarden een belangrijk instrument zijn om grip op AI te krijgen en in de toekomst te voldoen aan de AI-verordening.

Ruim een jaar na de inwerkingtreding van de AI-verordening zien we verdere verduidelijking en concretisering. De richtsnoeren van de Europese Commissie over verboden AI-systemen en de definitie van AI-systemen geven een eerste inzicht en uitleg, maar roepen ook veel vervolgvragen op. Verdere concretisering zal volgen, waarbij ook de ontwikkeling van standaarden duidelijkheid zal bieden over de manier waarop organisaties kunnen voldoen aan de eisen van de AI-verordening. Niet enkel generieke, maar ook sectorspecifieke doorvertalingen van de AI-verordening moeten leiden tot guidance en houvast.

10. Algoritmes opnemen in een register is een goed begin om de risico's die met algoritme-inzet gepaard gaan te beheersen. Een algoritmeregister bevat informatie over de inzet van algoritmes.

In grote lijnen heeft algoritmeregistratie twee overkoepelende doelen. Doel 1: Het bevorderen van interne controle op algoritmes (governance). Doel 2: Het bevorderen van externe controle op algoritmes (transparantie). De Nederlandse overheid werkt aan een wettelijk kader voor het Algoritmeregister. Het wettelijk kader moet zorgen voor duidelijkheid over het verplicht registreren van algoritmes. Ook zonder verplichting is het Algoritmeregister nu al een waardevolle en praktische tool voor het bieden van transparantie en het bevorderen van controle op overheidsalgoritmes. Nog lang niet alle organisaties benutten het Algoritmeregister; de AP moedigt overheidsorganisaties aan om hier zoveel mogelijk gebruik van te maken. In de bijlage 'Aan de slag met algoritme-registratie' geeft de AP acht concrete handvatten om aan de slag te gaan met algoritmeregistratie.

Overkoepelend beheersingsbeeld AI en algoritmes in Nederland – Zomer 2025

Beheersingspijler	Status Winter 2024-2025	Status Zomer 2025	Toelichting
Grip op ontwikkeling en volatiliteit van algoritmische- en AI-technologie	**Vraagt verhoogde aandacht**	**Vraagt verhoogde aandacht**	Ongewijzigd ten opzichte van het beheersingsbeeld 6 maanden geleden. De ontwikkelingen van (generatieve) AI gaan onverminderd door. De risico's en impact van frontier-modellen worden nog niet volledig begrepen.
Begrip en beheersbaarheid van nieuwe algoritmische en AI risico's	**Vraagt verhoogde aandacht**	**Vraagt verhoogde aandacht**	Ongewijzigd ten opzichte van het beheersingsbeeld 6 maanden geleden. Nieuwe ontwikkelingen zorgen voor nieuwe incidenten die niet altijd voorzien worden. Beheersbaarheids-instrumenten nog in ontwikkeling.
Ontwikkeling nationaal AI-ecosysteem	**Vraagt aandacht**	**Vraagt aandacht**	Het belang van een sterk nationaal AI-ecosysteem neemt toe. Dat zien we aan de groeiende wens voor digitale soevereiniteit in Nederland en Europa. De aandacht voor investeringen in AI neemt tevens toe, wat een inhaalslag mogelijk maakt.
Vertrouwen in, aandacht voor en kennis over algoritmes en AI in Nederlandse samenleving	**Ligt op koers**	**Ligt op koers**	Ongewijzigd ten opzichte van het beheersingsbeeld 6 maanden geleden. Na een dieptepunt in 2024 neemt het vertrouwen onder de Nederlandse bevolking in AI en algoritmes verder toe. Nederland is bewust bezig met AI-geletterdheid.
Kaders en bevoegdheden voor toezicht op AI-systemen	**Ligt op koers**	**Vraagt aandacht**	Vraagt meer aandacht vanwege de ambitieuze tijdslijnen voor de AI-verordening, de roep om simplificatie en de Nederlandse politieke situatie.
Geharmoniseerde en praktisch toepasbare standaarden voor AI-systemen	**Voortgang onvoldoende**	**Vraagt verhoogde aandacht**	Verbetering ten opzichte van het beheersingsbeeld 6 maanden geleden. Geharmoniseerde standaarden zullen op korte termijn nog niet gereed zijn om voor te bereiden op AI-verordening. Maar de voortgang in de ontwikkeling van standaarden begint te versnellen

Beheersingspijler	Status Winter 2024-2025	Status Zomer 2025	Toelichting
Registratie en transparantie over algoritmes en AI-systemen	**Vraagt verhoogde aandacht**	**Voortgang onvoldoende**	Beeld is minder positief ten opzichte van het beheersingsbeeld 6 maanden geleden. Er wordt al meerdere jaren aan algoritme-registratie gewerkt, maar er is nog een lange weg te gaan. Binnen de overheid hebben de meeste organisaties nog niets geregistreerd.
Zicht op incidenten bij inzet algoritmes en AI en borging van lessen	**Voortgang onvoldoende**	**Voortgang onvoldoende**	Ongewijzigd ten opzichte van het beheersingsbeeld 6 maanden geleden. Het ontbreekt aan meldingen en rapportages over incidenten met AI en algoritmes richting toezichthouders.
Institutionalisering van governance, risico-beheersing en auditering van algoritmes en AI	**Vraagt verhoogde aandacht**	**Vraagt verhoogde aandacht**	Ongewijzigd ten opzichte van het beheersingsbeeld 6 maanden geleden. Periodieke audits worden nog maar beperkt uitgevoerd en de uitkomsten zijn nog maar beperkt zichtbaar.

Toelichting: Als coördinerend toezichthouder op algoritmes en AI signaleert en analyseert de Autoriteit Persoonsgegevens (AP) proactief en overkoepelend de belangrijkste, sectoroverstijgende risico's en effecten. De zogenoemde beheersingspijlers geven inzicht in de verantwoorde omgang met deze risico's. Het overkoepelend beheersingsbeeld laat de actuele beheersing van algoritmes en AI in Nederland zien. Dit beheersingsbeeld moet worden gezien tegen de achtergrond van een brede maatschappelijke transitie, aangejaagd door AI als systeemtechnologie, die ieder jaar hogere eisen stelt aan het beheersniveau. Om de voortgang (mede ten opzichte van het vorige beheersingsbeeld) te duiden, hanteert de AP een kleurcodering per pijler: groen betekent 'ligt op koers', lichtroze betekent 'vraagt aandacht', paars betekent 'vraagt verhoogde aandacht' en donkerpaars betekent 'voortgang onvoldoende'.

1. Overkoepelende ontwikkelingen

SNEL NAAR DIT ONDERDEEL

1.1 Algoritmes als oplossing voor schaarste

In Nederland worden AI en algoritmes vaak ingezet met de hoge verwachting om schaarste in de maatschappij (voor een deel) op te lossen. Van AI en algoritmes wordt verwacht dat ze een slimme oplossing kunnen bieden voor bijvoorbeeld de tekorten op de woningmarkt en de druk op het energienetwerk. AI en algoritmes doen naar verwachting nieuwe suggesties of accurate voorspellingen waar organisaties op in kunnen spelen.

Zo worden er in de zorg verschillende algoritmes gebruikt om de werkdruk te verminderen of zorgtaken effectiever in te richten. Ook de regering zet vol in op de adoptie van AI in het zorgdomein.[1] De komende jaren investeert de overheid 400 miljoen euro in technologische innovaties, waaronder AI, om het werk in de zorg efficiënter te maken en de werkdruk te verlagen.[2] Zo werken een aantal ziekenhuizen samen in een pilot voor AI-toepassingen voor bijvoorbeeld detectie van fracturen en identificatie van incidentele longembolieën.[3] En wordt er bijvoorbeeld gewerkt met generatieve AI om patiëntvragen te helpen beantwoorden.[4]

De AP ziet dat AI in de zorg van grote betekenis kan zijn en voordelen oplevert voor zorgverleners en patiënten. Tegelijkertijd bestaat bijvoorbeeld het risico dat zorginstellingen afhankelijker worden van cloud-gebaseerde oplossingen en toenemende dataverwerking van gevoelige gegevens. Voor gebruik van generatieve AI komt daarbij dat veel toepassingen nog maar kort op de markt zijn en dat de impact op patiëntenzorg nog (deels) onbekend is. Het is niet voor niets dat de Inspectie Gezondheidzorg en Jeugd zorgaanbieders oproept om zorgvuldig om te gaan met aanschaf, invoering en gebruik van generatieve AI in de zorg.[5]

De kwaliteit van medische zorg kan ook toenemen, bijvoorbeeld via algoritmes die eerder longkanker op kunnen sporen… Regelmatig halen dit soort positieve ontwikkelingen het nieuws. Bijvoorbeeld een algoritme dat op basis van patronen in patiëntendossiers een eerdere voorspelling kan doen van longkanker.[6] Het algoritme analyseert patiëntendossiers op relevante patronen en wegen bepaalde risico's op longkanker. Vroege opsporing van longkanker met het nieuwe algoritme kan in sommige gevallen leiden tot betere behandeling van de ziekte.

…maar het gebruik van algoritmes en AI in de zorg kan leiden tot indirecte effecten. De mogelijkheden om AI in te zetten voor vroege opsporing van ziektes heeft ook effecten op andere delen van de zorgketen. De Raad voor Volksgezondheid en Samenleving (RVS) schrijft in een recent advies dat de zogenoemde diagnose-expansie (snelle toename van mensen met een ziektediagnose) zorgt voor meer druk op een al overbelast zorgstelsel.[7] Waar een vroege diagnose nuttig kan zijn voor mensen met klachten, zorgt een vroege diagnose bij mensen zonder klachten steeds meer voor oprekking van de criteria voor ziekte. De RVS roept daarom op tot een maatschappelijke discussie over het nut en de noodzaak van vroegdiagnoses nu de mogelijkheden daarvoor toenemen dankzij technologie zoals AI.

Ook het tekort op de woningmarkt wordt ondervangen door woningzoekenden via AI en algoritmes te ondersteunen. Zo worden AI-agents ingezet om woningzoekenden effectiever te laten zoeken naar een nieuwe woonplek, te helpen met bemiddeling, zoekresultaten te personaliseren en woningen te tonen waar zij een realistische kans op hebben.[8] Ook woningcorporaties verkennen de mogelijkheden van AI.[9]

De energiesector zet steeds vaker algoritmes in om overbelasting te voorkomen en zo de energietransitie te versnellen. Als vol wordt ingezet op AI, dan kan dat volgens het *International Energy Agency* 95 miljard euro op jaarbasis schelen en kan 175 GW aan energiecapaciteit worden vrijgespeeld.[10] In Nederland zetten energiebedrijven en netbeheerders vol in op AI, bijvoorbeeld om geschikte locaties voor zon- en windmolenparken te vinden of om energieverbruik te voorspellen voor optimalisatie. Toezichthouders ACM en AFM stelden vorig jaar vast dat algoritmische handel in energie sterk is toegenomen en wijzen op risico's van volatiliteit en ondoorzichtige prijsvorming.[11] Onderzoekers van het Rathenau Instituut waarschuwden er recent voor dat de integratie van AI in elektriciteitssystemen afhankelijkheden kan creëren van grote techbedrijven buiten de EU.[12] Daardoor worden netbeheerders en nutsbedrijven afhankelijk, neemt het democratisch toezicht af en ontstaan mogelijk monopolistische marktstructuren.

Mensen worden direct geraakt door de AI-gestuurde energiemarkt. Ontwikkelingen zoals dynamische prijsstelling, waarin AI een centrale rol speelt, stimuleren huishoudens om hun energieverbruik aan te passen aan schommelende tarieven. Dit kan helpen om piekbelasting te verminderen, maar roept ook vragen op over toegankelijkheid, eerlijkheid en uitlegbaarheid – vooral voor mensen die minder flexibel zijn of geen toegang hebben tot slimme technologieën.

AI biedt voordelen en kan bijdragen vraag en aanbod beter op elkaar af te stemmen of te sturen, maar roept ook nieuwe en fundamentele verdelingsvraagstukken op waarin op de achtergrond in feite afwegingen worden gemaakt over grondrechten en publieke belangen, zonder dat dat de volle aandacht krijgt. Wie krijgt wanneer toegang tot energie? Op basis waarvan worden keuzes gemaakt? En wie heeft daar invloed op? Daarbij is *algorithmic fairness* cruciaal. Maar wat 'eerlijk' is, hangt af van de norm: gaat het om gelijke kansen, gelijke behandeling of gelijke uitkomsten? Zonder duidelijke keuzes daarover kunnen algoritmes bestaande ongelijkheden juist versterken.[13]

1.2 Strategische autonomie en digitale soevereiniteit

Met toenemende geopolitieke spanningen, neemt de behoefte aan strategische autonomie verder toe. De integratie van AI in maatschappelijk-relevante sectoren en toepassingen vergroot afhankelijkheden. Strategische autonomie houdt in dat landen de mogelijkheid hebben om autonoom te handelen, zonder daarmee afhankelijk te zijn van andere landen.[14] Dit gaat onder meer om de Europese economie, energie, maar ook technologie.

Meta heeft in juni 2025 14,3 miljard dollar geïnvesteerd in het wereldwijde AI-trainingsbedrijf Scale. Met deze investering krijgt Meta 49% van het bedrijf in handen.[15] Ook is de CEO van Scale in dienst bij Meta gekomen. Investeringen op deze schaal geven aan hoeveel belang een kleine groep van grote techbedrijven hechten aan het ontwikkelen en trainen van AI-modellen.

Er zijn positieve ontwikkelingen die de digitale soevereiniteit en strategische autonomie vergroten. Toch is er nog een lange weg te gaan. Steeds vaker worden eigen systemen ontwikkeld om minder afhankelijk te worden van partijen buiten de EU. Ook zien we steeds vaker '*private AI stacks*', waar bedrijven en andere organisaties hun AI-toepassingen op kunnen bouwen en draaien. Maar volledige digitale soevereiniteit is nog ver weg. De Algemene Rekenkamer stelt namelijk dat de Rijksoverheid meer dan de helft van public-clouddiensten afneemt bij drie grote Amerikaanse techbedrijven.[16]

Verder zetten de provincie Groningen en enkele gemeenten in op de komst van een 'AI-factory', om de afhankelijkheid van andere landen op het gebied van kunstmatige intelligentie te verkleinen. De Europese Commissie wil minstens vijftien van deze 'fabrieken' in Europa hebben staan.[17] Zo'n fabriek moet bestaan uit een expertisecentrum en een supercomputer. Het wordt gezien als unieke kans voor economische groei, kennis-ontwikkeling en innovatie in de regio en in Nederland.[18] De fabriek gaat in totaal tussen de 160 en 240 miljoen euro kosten en wordt daarmee een van de duurste AI-fabrieken in Europa. Door financiele toezeggingen van regio Groningen, Noord-Drenthe en het Rijk komt de realisatie dichterbij.[19]

Werken aan strategische autonomie draagt uiteindelijk ook bij aan het behoud van grip over de algoritmes en AI die in Nederland gebruikt worden. In toenemende mate is de uitvoering van maatschappij-kritische processen, bijvoorbeeld overheidsdienstverlening, afhankelijk van de inzet van algoritmische processen. Strategische autonomie zorgt dat Nederland beschikt over eigen infrastructuur en eigen databeheer op basis waarvan algoritmes en AI-systemen kunnen opereren. Voor bijvoorbeeld algoritmes die worden ingezet ten behoeve van besluitvorming door overheid zorgt dit Nederland haar systemen volledig kan afstemmen op de eigen fundamentele waarden en daar de eigen technische en ethische controle over behoudt.

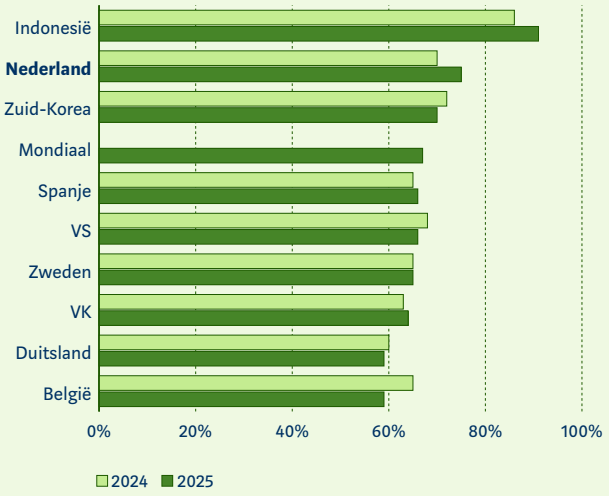
1.3 Vertrouwen, gebruik en zicht op incidenten

Nederlandse burgers worden steeds positiever over hun eigen AI-kennis en zien verhoudingsgewijs steeds meer voordelen in AI. Wereldwijd vindt ongeveer twee op de drie mensen dat ze een goed begrip hebben van AI. Dit blijkt uit internationaal vergelijkend onderzoek van Ipsos in 30 landen.[20] In Nederland is inmiddels 75% van de burgers het eens met deze stelling – een toename van 5% ten opzichte van 2024. Daarmee scoort Nederland hoger dan de meeste andere landen. Het vertrouwen in de eigen AI-kennis is dus goed. Ook zien Nederlanders verhoudings-gewijs steeds meer voordelen dan nadelen in AI. In 2024 waren Nederlanders wereldwijd het meest kritisch over AI. 36% van de Nederlandse bevolking gaf toen aan meer voordelen dan nadelen te zien in AI. In 2025 is dit percentage met 7% gestegen naar 43%. Daarmee is Nederland op een vergelijkbaar niveau gekomen met landen als België, de Verenigde Staten (VS), het Verenigd Koninkrijk (VK) en Zweden. Deze ontwikkeling bevestigt het beeld dat we constateerden in de vorige editie van de RAN: de neerwaartse trend in het Nederlandse vertrouwen in algoritmes en AI is gekeerd. Zie ook figuur 1.1.

FIGUUR 1.1 | ONTWIKKELINGEN IN PERCEPTIE OVER EN GEBRUIK VAN AI IN NEDERLAND

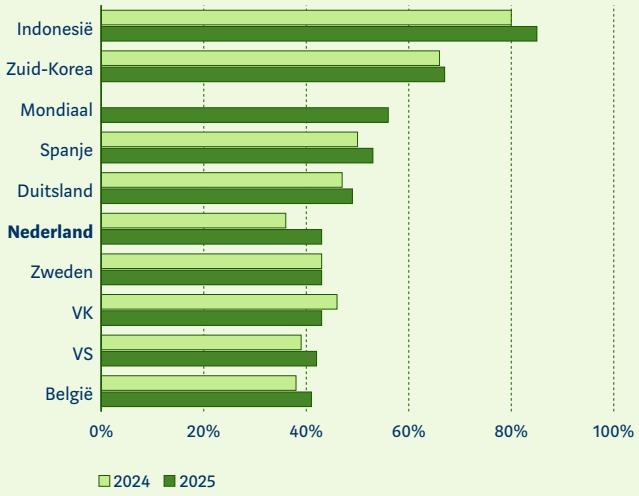
In 2025 zijn Nederlandse burgers zeer positief over hun begrip van AI...

% Eens met de stelling: "Ik heb een goed begrip van wat AI is"



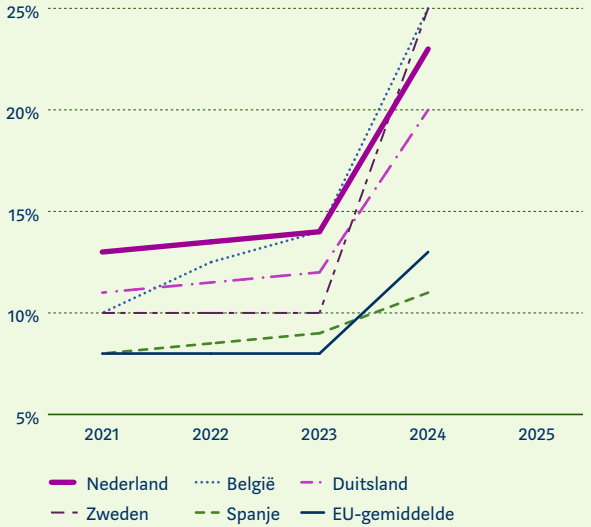
... staan Nederlanders relatief minder kritisch tegenover AI...

% Eens met de stelling: "Producenten en diensten met AI hebben meer voordelen dan nadelen"



... en met AI-gebruik door bedrijven blijft Nederland een koploper.

% Bedrijven dat gebruik maakt van minstens één van zeven AI-technologiën



Bron: links en midden: Ipsos AI monitor (n=23.685, 32 landen), rechts: Eurostat (online data code: isoc_eb_ain2)

Steeds meer Nederlandse bedrijven gebruiken AI, met een versnelling in de afgelopen jaren. Eurostat-data geeft inzicht in de ontwikkeling in de afgelopen tijd. In 2024 gebruikte ongeveer één op de vier Nederlandse bedrijven minimaal één van de zeven AI-technologieën die door Eurostat zijn gedefinieerd (figuur 1.1). Het gaat daarbij om machine learning, afbeeldingsherkenning, robotica in autonome voertuigen, robotica bij proces-automatisering, spraakherkenning, tekst-mining en taalgeneratie (geschreven of gesproken). Met het percentage van 23% blijft Nederland een Europese koploper. Maar we zijn qua AI-gebruik binnen bedrijven tegelijkertijd wel ingehaald door bijvoorbeeld België en Zweden.

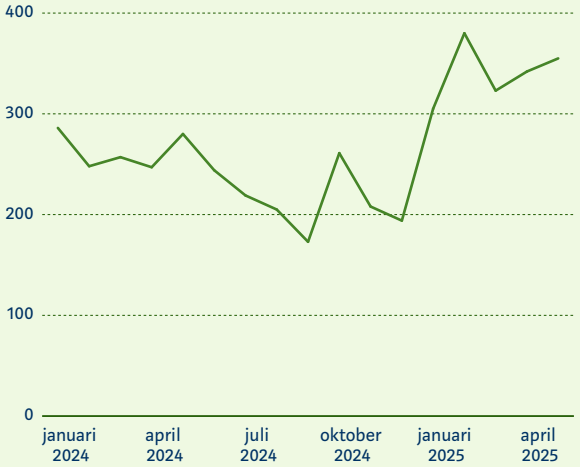
De afgelopen maanden doet de media wereldwijd steeds vaker bericht van incidenten en risicovolle ontwikkelingen rond AI. Op basis van de informatie uit de media worden sinds begin 2025 elke maand zo'n 300 tot 400 incidenten en risicovolle ontwikkelingen toegevoegd aan de OECD AI Incidents Monitor. Dit help om zicht te krijgen in risicovolle ontwikkelingen en de manier waarop zij aan principes voor verantwoorde AI raken. In 2024 ging het om 200 tot 300 incidenten per maand. Figuur 1.2 toont de recente ontwikkeling.

Daarbij valt op dat de incidenten en risicovolle ontwikkeling aan alle principes voor verantwoorde AI raken. Meldingen met betrekking tot de verantwoorde-AI-principes, rechtvaardigheid (fairness) en duurzaamheid komen het minst voor. Dat heeft er deels mee te maken dat deze risico's bij rechtvaardigheid nog veelal onder de radar blijven of zich bij duurzaamheid primair op de lange termijn afspelen. Over de periode januari tot mei 2025 deden ongeveer 700 incidenten en risicovolle ontwikkelingen zich voor op het gebied van (gebrek aan) transparantie (figuur 1.2). Maar ook op andere terreinen zoals digitale veiligheid (waaronder cybersecurity-risico's), verantwoording (accountability), privacy, grondrechten en fysieke veiligheid zijn het aantal incidenten en risicovolle ontwikkelingen talrijk.

Incidenten kunnen impact hebben op rechten én de fysieke wereld. Zo ontstond een privacy-incident doordat prompts van sommige gebruikers in Meta AI, dat binnenkort ook in Nederland beschikbaar is, op een openbare feed van Meta verschenen zonder toestemming van de gebruiker. Dat kan bijvoorbeeld gebeuren bij gebruikers die Meta AI via Instagram gebruiken en een openbaar Instagramaccount hebben.[21] Verder creëerde Google Maps onlangs onnodige risico's voor de fysieke veiligheid: de AI van het programma gaf op dag 1 van het Hemelvaartsweekend onterecht aan dat diverse snelwegen in Duitsland, Nederland en België gesloten zouden zijn, waardoor alternatieve wegen overvol raakten.[22] Gemelde cybersecurityrisico's ontstaan veelal doordat AI hacken via bijvoorbeeld social engineering en phishing een stuk makkelijker maakt.

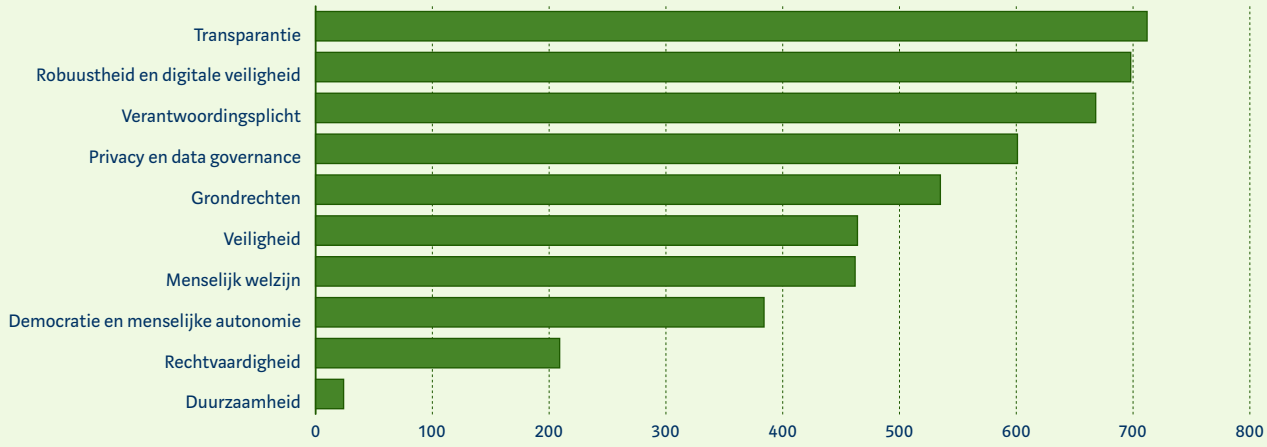
FIGUUR 1.2 | WERELDWIJDE ONTWIKKELINGEN IN AI-INCIDENTEN EN RISICOVOLLE AI-ONTWIKKELINGEN

Wereldwijd is er een gestage toename van het maandelijkse aantal geïdentificeerde AI-incidenten...



Geïdentificeerde AI-incidenten en risicovolle ontwikkelingen per maand (wereldwijd)

... en deze raken in de praktijk aan alle principes voor verantwoorde AI.



Aantal incidenten en risicovolle ontwikkelingen per principe voor verantwoorde AI (wereldwijd, januari 2025 - mei 2025, cumulatief)

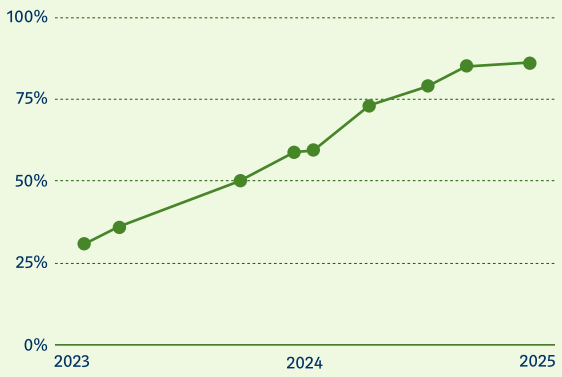
Bron: OECD AI Incidents Monitor (AIM)

1.4 Ontwikkeling generatieve AI gaat onverminderd voort

Nieuwe generatieve AI-modellen worden ook in 2025 weer met grote regelmaat gelanceerd. Grote aanbieders van taalmodellen (zogenoemde *Large Language Models*), zoals OpenAI, Google, Meta, Mistral en nieuwkomers als DeepSeek en Alibaba, hebben in de eerste helft van dit jaar gezamenlijk al meer dan 10 nieuwe (versies van) modellen uitgebracht. Ondanks dat deze modellen op een aantal gebieden tegen hun grenzen lijken aan te lopen wordt nog steeds belangrijke vooruitgang geboekt in de bruikbaarheid en het economische nut van deze modellen.

Nieuwe taalmodellen zijn de laatste tijd vaak niet veel 'slimmer' meer dan hun voorgangers, maar verbeteren prestaties op andere vlakken. Op een belangrijke graadmeter voor de prestaties van modellen, ook wel de GPQA-score, lijkt het afgelopen half jaar een plateau te zijn bereikt.[23] (Zie figuur 1.3). Maar er wordt nog steeds belangrijke vooruitgang geboekt op het gebied van (i) de typen content die een model kan genereren, zoals tekst en video, (ii) de hoeveelheid input waar een model mee uit de voeten kan, en (iii) de efficiëntie van modellen.

FIGUUR 1.3 | DE GROEI VAN TAALMODELLEN VLAKT OP SOMMIGE TERREINEN AF



Toelichting: Hoogste score van LLM op GPQA-index (% goede antwoorden, maximum is 100%).
Bron: LLM Stats

Generatieve AI-modellen kunnen met steeds meer input overweg. Waar bij de lancering van ChatGPT eind 2022 de zogenoemde *context window size*, de omvang van een prompt of document die een model als input krijgt, amper één bladzijde uit een boek was, kunnen de nieuwste modellen overweg met ruim twee volle boeken aan input.[24] De nieuwste generatieve AI-modellen gebruiken en genereren naast tekst inmiddels ook code, audio en video. Deze modellen kunnen bijvoorbeeld in tekst samenvatten wat er in een video gebeurt, of een videoclip met geluid genereren op basis van een *prompt* van een gebruiker.

Modellen halen vergelijkbare resultaten met minder rekenkracht. Als gevolg van voortdurende innovaties in de manier waarop modellen worden getraind en werken, behalen nieuwe modellen vaak vergelijkbare prestaties als eerdere topmodellen, tegen een fractie van de kosten.[25] Deze efficiëntere modellen gebruiken minder rekenkracht en daarmee energie, zowel tijdens de ontwikkeling en bij het gebruik.

Tegelijkertijd blijven sommige modeltekortkomingen hardnekkig, zoals wanneer modellen foute informatie fabriceren. Veel modellen beschikken tegenwoordig over toegang tot het internet voor het vinden van informatie en komen tot hun output op basis van een aantal 'redeneer-stappen'. Hoewel deze functionaliteiten in bepaalde situaties tot betere uitkomsten leiden, laten verschillende studies zien dat het probleem blijft bestaan dat modellen foute informatie uit het luchtledige blijven fabriceren.

Organisaties in Californië zijn onvoldoende in staat om risico's en impact van frontier-modellen volledig te begrijpen. Op 17 juni verscheen 'The California Report on Frontier AI Policy'.[26] Dit rapport stelt dat er een goede balans nodig is tussen innovatie en regulering. Maar dat organisaties nu nog niet in staat zijn om de snelle ontwikkelingen bij te blijven waar het gaat om risico-management en potentiële impact. Het rapport adviseert hoe beleidsinstrumenten gebruikt kunnen worden om belangrijke principes te waarborgen, in lijn met de EU en het VK. Deze inzet wijkt af van de voorstellen op het federale niveau in de Verenigde Staten, waar juist een verbod op verdere regulering van AI wordt besproken.[27]

Personen die teksten schrijven met behulp van een LLM/AI-chatbot gebruiken hun brein minder, slaan de informatie uit die teksten amper of niet op, zijn minder creatief en de uiteindelijke tekst gebruikt vooral

gangbare woorden en concepten. Dit blijkt uit de eerste uitkomsten van een onderzoek dat MIT heeft uitgevoerd.[28] In dit onderzoek werd een groep van 54 deelnemers gevraagd om een essay te schrijven zonder tools, met gebruik van een traditionele zoekmachine, en met behulp van ChatGPT. Deze eerste studie maakt deel uit van het in kaart brengen van de cognitieve effecten van het gebruik van AI-tools voor het schrijven in de onderwijscontext. Deze eerste signalen zijn reden tot zorg en laten zien dat extra aandacht nodig is voor goed doordachte integratie van deze tools in het onderwijs en de samenleving.

Een voorbeeld van toenemende inzet is het gebruik van generatieve AI in chatbots voor publiekscommunicatie. Dit soort chatbots maken steeds vaker gebruik van grote taalmodellen, waarbij deze vervolgens zijn afgesteld op de specifieke informatie die een organisatie in haar publiekscommunicatie wil verstrekken. De communicatie van de chatbot wordt op deze manier een samenspel tussen de kenmerken van het onderliggende taalmodel en de specifieke informatie en instructies van de organisatie. De informatieverstrekking door de organisatie wordt daarmee mede beïnvloed door ontwikkelingen, zoals updates, in het onderliggende taalmodel. Ook is het moeilijk om voor alle omstandigheden voor te schrijven hoe de chatbot moet communiceren. Dit brengt nieuwe uitdagingen. In de komende periode gaat AP verder kijken naar de inzet van generatieve AI voor publiekscommunicatie.

1.5 AI en energieverbruik

De enorme hoeveelheden data en rekenkracht kan problemen opleveren voor het belang van een schone, gezonde en duurzame omgeving.[29] Het kost steeds meer energie om nieuwe systemen te trainen, wat dit belang steeds verder onder druk zet. Het energieverbruik houdt niet op na het trainen van een AI; het gebruik van AI vergt op zijn minst net zo veel energie. Wanneer een AI-systeem getraind is, betekent het niet direct dat het energieverbruik afneemt. Zo kost het stellen van een vraag aan ChatGPT naar schatting 10 keer meer energie dan dezelfde vraag stellen aan Google.[30]

Grote AI-bedrijven zetten vol in op eigen (nucleaire) energievoorziening. Het World Economic Forum stelt dat de toename van het energieverbruik vraagt om een versnelling van de energietransitie, waarbij we gebruik moeten maken van duurzame oplossingen.[31] Om in de energiebehoefte te voorzien, werken grote AI-bedrijven aan eigen infrastructuur en voorzieningen. Zo is er bijvoorbeeld meer interesse in nucleaire energie.[32] Microsoft heeft een overeenkomst van 1,6 miljard dollar aangekondigd om voor het gebruik van AI stroom op te wekken met een kernreactor, terwijl Google en Amazon ook overeenkomsten op het gebied van nucleaire energie hebben om hun AI-verrichtingen te ondersteunen.[33] Naast de grote impact op milieu en duurzaamheid brengt de ontwikkeling van AI ook het risico met zich mee dat essentiële (energie-)infrastructuur in private handen terechtkomt, waardoor de maatschappelijke en individuele kosten toenemen.

Hoewel precieze cijfers ontbreken, neemt ook het watergebruik voor AI hard toe. Datacentra gebruiken grote hoeveelheden water om de hardware voor AI te koelen en oververhitting te voorkomen. Microsoft beheert een groot deel van de serverinfrastructuur waarop onder andere ChatGPT draait. Sinds de komst van ChatGPT in 2022 schatten onderzoekers dat het waterverbruik van Microsoft met 34% steeg in 2022 ten opzichte van een jaar daarvoor. Eén e-mail van slechts 100 woorden laten schrijven door GPT-4 kost 519 milliliter water.[34]

Vanwege het grote waterverbruik van datacenters kan er waterschaarste ontstaan in gebieden die over minder water beschikken.[35][36] In de VS staan datacenters vaak in gebieden waar al een bepaalde mate van waterschaarste heerst.[37] Het is belangrijk dat er meer nadruk komt te liggen op hernieuwbare energiebronnen en duurzame AI-systemen. Groene AI focust zich enerzijds op het inzetten van AI om duurzaamheidsvraagstukken op te lossen en anderzijds op het optimaliseren van AI zodat het minder veeleisend wordt.[38]

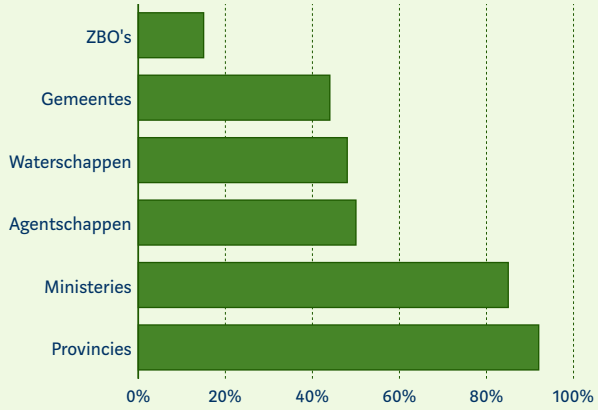
1.6 Algoritmes en AI-systemen in kaart

Het vullen van het Algoritmeregister voor de Nederlandse overheid vordert gestaag, maar de meeste overheidsorganisaties hebben nog altijd geen algoritmes geregistreerd. Sinds de start van het Algoritmeregister voor de Nederlandse overheid in 2023 zijn er inmiddels ongeveer 1.000 algoritmes geregistreerd. Koplopers zijn de gemeente Amsterdam met ongeveer 60 algoritmes en de Douane met ongeveer 50. Tegelijkertijd hebben de meeste overheidsorganisaties nog altijd geen enkel algoritme geregistreerd. Zo heeft minder dan één op de vijf zelfstandige bestuursorganen minimaal één algoritme in het register opgenomen. Ook meer dan de helft van alle gemeentes heeft nog niets geregistreerd. Dit laat figuur 1.4 zien.

FIGUUR 1.4 | ONTWIKKELING ALGORITMEREGISTRATIE NEDERLANDSE OVERHEID

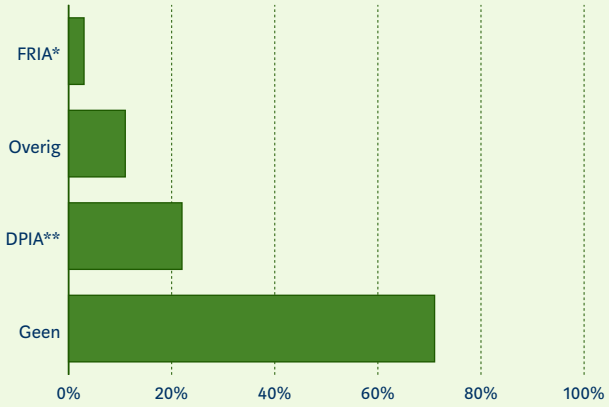
De meeste overheidsorganen hebben nog geen algoritmes geregistreerd...

% Organisaties met minimaal één registratie in algoritmeregister



... en voor de meeste algoritmes is nog geen impacttoets uitgevoerd.

% Algoritmes waarvoor wel of geen impacttoets is uitgevoerd



* Fundamental rights impact assessment (grondrechtenbeoordeling)
** Data protection impactassessment (gegevensbeschermingseffectbeoordeling)

Bron: Algoritmeregister (algoritmes.overheid.nl), peildatum 11 juni 2025

Hoewel het mogelijk is dat organisaties geen enkel algoritme gebruiken, is dit in veel gevallen onwaarschijnlijk... De kennisopbouw over algoritmes en AI binnen organisaties is nog in volle gang. En algoritmes en AI-systemen zijn niet altijd makkelijk als zodanig te herkennen. Het is belangrijk dat organisaties hierbij worden geholpen en dat zij extern uitgedaagd kunnen worden op hun algoritmeregistratie en het eventuele gebrek daaraan. Hoe dan ook, het blijft lastig om op dit moment in te schatten hoever organisaties zijn met hun algoritmeregistratie. Zo concludeert de Wetenschappelijke Adviesraad Politie dat er op het gebied van data en AI veel gebeurt binnen en buiten de politieorganisatie.[39] Hierbij constateert de adviesraad dat "een volledig overzicht hiervan lastig te geven is, mede omdat de ontwikkelingen snel gaan en er weinig afstemming is tussen de talrijke initiatieven op verschillende plekken binnen de organisatie". Het is met dat in het achterhoofd lastig te duiden hoe dit zich verhoudt tot het huidig aantal van zes geregistreerde algoritmes door de Nationale Politie (zie ook box 1.1 over algoritmes en AI bij de Politie).

... Het is daarom belangrijk, als deze situatie zich voordoet, óók te registeren dat een overheidsorganisatie verklaart geen algoritmes te gebruiken. Binnen de Rijksoverheid kan men dit bijvoorbeeld baseren op de brieven die per ministerie aan de Tweede Kamer worden gestuurd over de voortgang van de algoritmeregistratie, waarin ook staat hoeveel algoritmes en AI-systemen binnen organisaties zijn geïdentificeerd.

Aandachtspunt in de algoritmeregistratie is ook dat overheidsorganisaties voor minder dan 5% van de algoritmes een Fundamental Rights Impact Assement (FRIA), de grondrechteneffectenbeoordeling, hebben uitgevoerd. Figuur 1.4 toont de mate waarin impacttoetsen zijn uitgevoerd voor algoritmes. Een FRIA kan op verschillende manieren via verschillende formats worden uitgevoerd. Bijvoorbeeld door het format van de *Impact Assessment Mensenrechten en Algoritmes* (IAMA) te volgen. Idealiter worden de uitkomsten van zo'n beoordeling ook openbaar, zodat de belanghebbenden zelf ook kunnen nalezen hoe de effecten op hun grondrechten zijn beoordeeld. In het algoritmeregister van de Nederlandse overheid is deze onderliggende documentatie tot nu toe nog niet of nauwelijks opgenomen.

Het uitvoeren van een FRIA wordt onder de AI-verordening een verplichting voor overheidsorganisaties die hoog-risico AI-systemen gebruiken... Elke overheidsorganisatie moet vóór de eerste inzet van een hoog-risico AI-systeem een FRIA uitvoeren. Het gaat hier om systemen die gebruikt worden bij biometrie, onderwijs, arbeid, publieke dienstverlening, rechtshandhaving, migratie,

asiel- en grensbewaking, rechtspraak of democratische processen. Tijdens het gebruik van zo'n hoog-risico systeem wordt deze beoordeling geactualiseerd als het nodig is.

... En overheidsorganisaties stellen de markttoezicht-houder vervolgens op de hoogte van de uitkomsten van de FRIA voordat ze het systeem inzetten. Het AI-office, onderdeel van de Europese Commissie voor Europees AI-toezicht en coördinatie, ontwikkelt de komende periode een sjabloon die gebruikt kan worden voor de FRIA. De rapportageverplichting over de FRIA gaat markt-toezichthouders helpen om actueel inzicht te krijgen in het gebruik van AI-systemen binnen de overheid en de bijbehorende inschatting van grondrechtenrisico's. Binnen de financiële sector gaat een soortgelijke verplichting gelden voor kredietverstrekking en premiestelling.

Box 1.1

Casus: Inzet van algoritmes en AI door de politie

De inzet van algoritmes bij de politie blijft aandacht vragen. Politiewerk wordt steeds mee gestuurd door data en de vraag is wat de nieuwe context betekent voor de uitvoering en inrichting van het politiewerk. In een rapport van De Wetenschappelijke Adviesraad Politie worden de verschillende uitdagingen van data- en AI-toepassingen bij de Politie nader belicht.[40] Besluitvorming over AI moet volgens het advies onderdeel worden van de bredere governance van de politieorganisatie. Ethische, juridische en maatschappelijke aspecten moeten structureel worden meegenomen bij het ontwikkelen van AI-toepassingen. Actieve transparantie moet de regel worden. AI-geletterdheid moet vorm krijgen als een 'integrale basisvaardigheid', ter ontwikkeling van een 'kritische digitale mindset', met oog voor (conflicterende) publieke waarden. Ook stelt de raad dat gekeken moet worden naar AI-toepassingen die de relatie van de Politie met burgers kunnen verbeteren, om zo mogelijk ook het maatschappelijke debat rondom de inzet van AI door de politie de positieve kant op te sturen. Er moet bovendien meer empirisch onderzoek komen naar de effectiviteit van inzet van AI door politie, en een normatieve toets op wenselijkheid en uitlegbaarheid. Tenslotte adviseert de raad om niet verder in te zetten op systemen die gedrag op persoonsniveau moeten voorspellen.

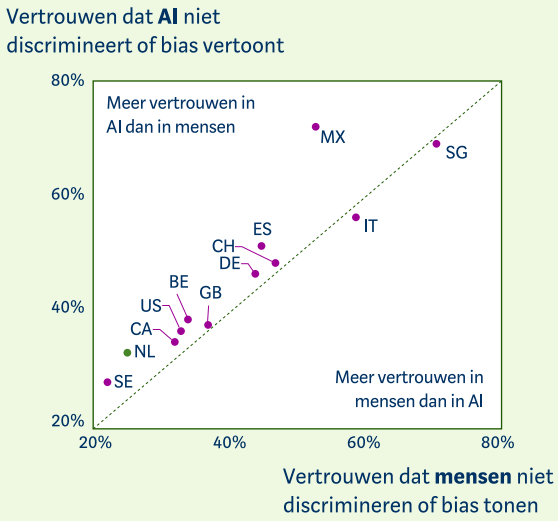
Bepaalde voorspellende AI-systemen voor de risicobeoordeling van strafbare feiten zijn verboden op basis van de AI-verordening. De AP wijst er op dat volgens artikel 5, lid 1, sub d van de AI-verordening ("verbod D") AI-systemen niet mogen worden gebruikt om iemands risico op strafbaar gedrag te beoordelen als dat alleen gebeurt op basis van profilering of persoonlijke eigenschappen. Dit sluit aan bij het advies van de Raad om geen AI-systemen in te zetten die individueel gedrag proberen te voorspellen. De AP heeft in februari een 'Oproep tot input' gelanceerd naar aanleiding van dit verbod. Daarbij is een toelichting gegeven op specifieke criteria voor deze verboden AI-systemen en vragen gesteld voor aanvullende input.

1.7 Discriminatierisico's bij algoritmes en de vormgeving van menselijke toetsing

In 2025 zijn Nederlanders zich zeer bewust dat niet alleen mensen maar ook algoritmes kunnen discrimineren richting groepen burgers... Nederlanders hebben, samen met Zweden en Canadezen, relatief weinig vertrouwen in het vermogen van AI om eerlijk te handelen zonder discriminatie of vooroordelen. Slechts 33% van de Nederlanders gelooft dat AI-systemen geen bias vertonen. In andere landen, bijvoorbeeld Duitsland (47%), Spanje (52%) en Mexico (73%) ligt dit percentage hoger. Opvallend is dat het vertrouwen in mensen op dit vlak nog lager ligt: slechts 25% van de Nederlanders denkt dat mensen zelf niet discrimineren of bevooroordeeld zijn. Er lijkt een duidelijke samenhang te zijn: in landen waar men weinig vertrouwen heeft in mensen op het gebied van non-discriminatie, is ook het vertrouwen in AI-systemen laag. Toch geldt in de meeste landen dat mensen AI nét iets eerlijker inschatten dan andere mensen. (Zie figuur 1.5).[41]

... Waarbij er onder Nederlanders ook een sterk zelfbewustzijn is dat mensen bias vertonen en (onbewust) discrimineren. Een onderzoek van de Europese Commissie uit 2023 toont dat Nederlanders zich verhoudingsgewijs sterk bewust zijn van hun eigen bias tegen anderen. In 2023 gaf 26% van de Nederlanders aan in de afgelopen 12 maanden weleens een ander bewust of onbewust gediscrimineerd te hebben. Dit is bijna vier keer zoveel als het Europees gemiddelde van 7%. Na Nederland worden de hoogste percentages behaald in Zweden (19%), Denemarken (16%) en Roemenië (11%). Alle overige lidstaten rapporteren onder de 10%.[42]

FIGUUR 1.5 | VERTROUWEN DAT AI EN MENSEN NIET DISCRIMINEREN OF BIAS VERTONEN



Toelichting: Nederlanders hebben weinig vertrouwen dat AI en mensen niet discrimineren.

Bron: Ipsos AI monitor (n=23.685, 32 landen)

Deze uitkomsten tonen het belang om, in processen die van invloed zijn op mensen, zowel algoritmische bias als menselijke bias terug te dringen. Het sterke bewustzijn dat beiden zich voordoen biedt voor Nederland een goede uitgangspositie en draagvlak om hier stappen in te zetten. Zeker het toeslagenschandaal heeft voor Nederlanders duidelijk gemaakt dat sprake kan zijn van institutioneel vooringenomen handelswijzen in onderzoek en beoordeling door een overheidsorganisatie.[43] Deze institutionele vooringenomenheid wordt verder versterkt als risicoselectie plaatsvindt aan de hand van discriminerende algoritmes, zoals ook bij het toeslagenschandaal aan de orde was.[44]

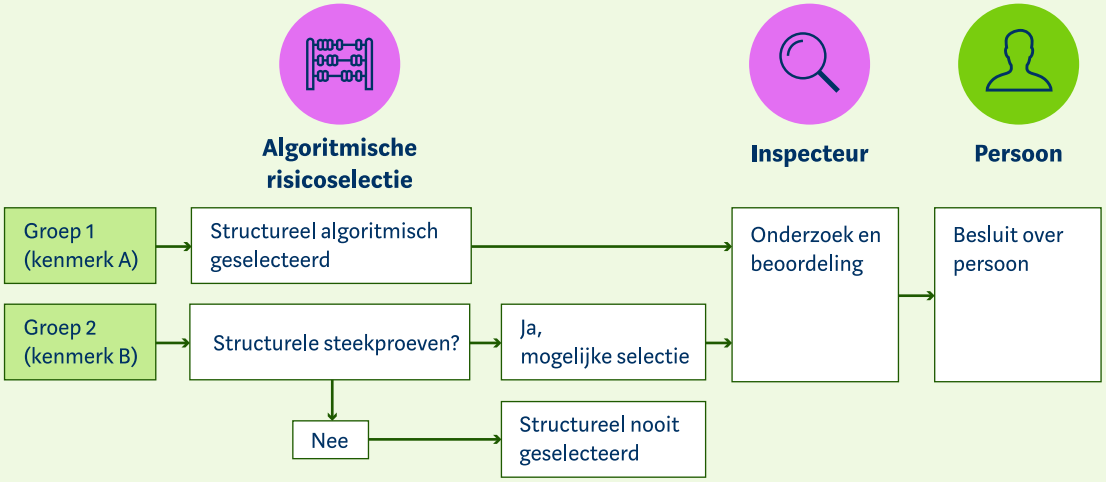
Om onder andere bias- en discriminatierisico's zoveel mogelijk tegen te gaan, wijst de wetenschap op het belang de wisselwerking tussen mens en algoritme goed vorm te geven. Bias door algoritmes zijn extra risicovol: ze werken systematisch, kunnen elke persoon op dezelfde manier treffen, leggen historische vooroordelen vast én zijn vaak lastig uit te leggen of te doorzien. Deze bias komen bovenop menselijke vooroordelen, al kan de vorm ervan verschillen.[45]

Wanneer verkeerd vormgegeven, kan de wisselwerking tussen algoritmes en mensen de mate van discriminatie en bias zelfs verergeren. Neem het voorbeeld van profilerende en selecterende algoritmes die een voorselectie doen voor fraudedetectie door inspecteurs. Verschillende studies bevestigen het risico dat inspecteurs, op basis van algoritmische aanbevelingen meer inconsistent en bevoordeeld kunnen handelen.[46] Tegelijkertijd kunnen algoritmes en mensen elkaar bij een goede vormgeving juist aanvullen. Een belangrijk doel is om inspecteurs en algoritmes zo op elkaar af te stemmen dat inspecteurs precies genoeg, maar geen blind vertrouwen hebben in de uitkomsten van een algoritme. Zodat inspecteur en algoritme complementair zijn aan elkaar. Dat vereist heldere, ondersteunende uitleg over hoe het algoritme zaken selecteert en risico's inschat. Denk bijvoorbeeld aan het tonen van onzekerheidsmarges bij elke risicoselectie, zodat inspecteurs beter kunnen beoordelen hoe stevig een signaal is.

In mei 2025 is in de Tweede Kamer een motie aangenomen waarin de Nederlandse regering wordt verzocht als uitgangspunt te nemen dat burgers "blind" beoordeeld worden, bijvoorbeeld bij fraudedetectie. De gedachte is hierbij dat als een inspecteur geen kennis heeft van de aanleiding waarom een individueel dossier onderzocht moet worden, de inspecteur ook niet beïnvloed kan worden door de algoritmische voorspelling. De motie heeft daarmee tot doel om tunnelvisie en vooringenomenheid in de menselijke beoordeling te voorkomen.

FIGUUR 1.6 | STRUCTURELE STEEKPROEVEN ZIJN NODIG OM DISCRIMINATIERISICO'S TE BEHEERSEN, OOK BIJ BLINDE CONTROLE

Structurele toevoeging van steekproeven is cruciaal om bias in de gehele keten te detecteren, ook als inspecteurs "blind" dossiers beoordelen.



Toelichting: In dit voorbeeld vertoont het algoritmische selectiemodel ongewenste discriminatie richting groep 1 (op basis van kenmerk A). Deze groep wordt structureel gecontroleerd, bijvoorbeeld in fraudedetectie, terwijl groep 2 nooit wordt geselecteerd door het algoritme. Hoewel automatiseringsbias kan worden voorkomen door inspecteurs niet te vertellen dat te onderzoeken dossiers zijn geselecteerd door een algoritme, kan het risico op discriminatie door de keten heen alleen worden teruggedrongen door de dossiers voor inspecteurs te verrijking met steekproeven die ook groep 2 omvatten. Op deze manier kunnen vals-negatieven aan het licht komen (dossiers die onderzocht zouden moeten worden, maar die niet waren geselecteerd door het algoritme). Dit is nodig om de accuraatheid van het selectie- algoritme gedurende de levenscyclus te waarborgen en het discriminatierisico te beheersen.

De AP vindt hierbij drie punten belangrijk: (i) bekijk ook de besluitvormingsketen als geheel, dit is noodzakelijk; (ii) voeg aselecte steekproeven aan de risicoselectie toe; en (iii) zorg dat de risicoselectie uitlegbaar is en aangevochten kan worden. De totstandkoming van een risicobeoordeling over een persoon, bijvoorbeeld tijdens fraudeonderzoek, is uiteindelijk terug te voeren op een combinatie van algoritmische (voor)selectie en de daaropvolgende menselijke controle. Blinde beoordeling van dossiers door inspecteurs kan op individueel niveau geen discriminatie en bias wegnemen die zich in het voorliggende stadium heeft voltrokken. Door het proces standaard te verrijken met een aselecte steekproef ontstaat er op overkoepelend niveau een basis voor periodieke audits. Deze audits kunnen dan regelmatig controleren of zich in het algoritme – en in het hele proces – bias of discriminatie voordoet (zie figuur 1.6). Daarmee kan bias en discriminatie uiteindelijk worden teruggedrongen. Als inspecteurs in eerste instantie een blinde beoordeling moeten doen, is het noodzakelijk om op basis van een aselecte steekproef dossiers toe te voegen. Anders weten inspecteurs op voorhand dat elk dossier al op basis van het algoritme is voorgeselecteerd en dus als "risicovol" is beoordeeld.

En ook bij een blinde beoordeling moeten inspecteurs wel in staat zijn om (i) het algoritme te kunnen controleren en (ii) burgers de mogelijkheid te geven de risicoselectie aan te vechten of de selectie aan hen uit te leggen. Enig inzicht voor de inspecteur in de uitkomsten van de algoritmische risicoselectie is dus nodig. Om het concreet te maken: zonder deze informatie kan een inspecteur, als deze bijvoorbeeld contact opneemt met een burger, niet uitleggen waarom een persoon is geselecteerd voor een onderzoek. Belangrijk is dat

inspecteurs, ook bij een blinde controle, op het juiste moment een begrijpelijke en bruikbare uitleg krijgen over de algoritmische risicoselectie. En de inspecteur moet voldoende AI-geletterd en bekwaam zijn om de wijze waarop het algoritme tot haar selectie komt te kunnen beoordelen. De volgorde waarin data aan een inspecteur wordt aangeboden is daarbij van belang, omdat dit een vervolgbeslissing kan sturen.[47] Dit kan uiteindelijk leiden tot de tunnelvisie en vooringenomenheid die in de Kamermotie aan de orde is gekomen. De AP adviseert om "blinde beoordeling" te zien als de volgorde waarin informatie, en dan vooral de risico-inschatting van een algoritme, aan een inspecteur wordt getoond. Blinde beoordeling houdt dan in dat inspecteurs het risico-oordeel van het algoritme pas te weten komen nadat deze eerst een eigen oordeel hebben gevormd.

Box 1.2

Casus: Facebook's advertentiealgoritme discrimineerde bij vacatures

Op 18 februari 2025 oordeelde het College voor de Rechten van de Mens dat de aanbevelingsalgoritmes van Facebook discrimineren op grond van geslacht.[48] Het College is bevoegd om in Nederland een oordeel te vellen bij schendingen van het gelijkebehandelingsrecht. De stichtingen Clara Wichmann en Global Witness klaagden bij het College over Facebook omdat uit onderzoek van Global Witness zelf bleek dat vacatures met stereotype mannen- en vrouwenberoepen vooral aan diezelfde groep getoond werden. Zo bleek bij een vacature voor de functie van receptionist dat deze in 2022 voor 96% en 2023 97% aan vrouwen werden getoond, terwijl een vacature voor monteur voor 96% in diezelfde jaren aan mannen werd getoond. Het algoritme kan hierdoor stereotypering bevorderen, want zo worden vrouwen die bijvoorbeeld automonteur willen worden belemmerd in hun zoektocht naar een passende baan. Het College vindt het onderzoek voldoende om een vermoeden van indirecte discriminatie aan te nemen.

Facebook is er niet in geslaagd om de discriminerende en stereotyperende effecten van het algoritme te weerleggen of te rechtvaardigen. Facebook heeft de mogelijkheid erkend dat geslacht meer weging in het advertentiealgoritme heeft gekregen door het like- en klikgedrag van gebruikers van het platform. Dat er geen verklaring is voor de onderzoeksresultaten van Global Witness, acht het College voor rekening van Facebook. Ook slaagt Facebook er niet in om een objectieve rechtvaardiging te geven voor het discriminerende onderscheid. Het College vindt dat Facebook als socialmediaplatform verantwoordelijk is om de werking van het algoritme goed te monitoren en onderzoek moet doen naar stereotypering als gevolg van de inzet van het advertentiealgoritme. Facebook heeft niet concreet kunnen – of willen – aantonen of, en zo ja, hoe Facebook concreet monitort en onderzoek hiernaar doet. Het oordeel is kenmerkend voor recente rechtelijke uitspraken waarbij geen uitleg gegeven kon worden over de werking en impact van een algoritme, zoals in de SyRI-zaak.[49] Gebrekkige transparantie werkt in de rechtszaal in het nadeel van degene die verantwoordelijk is voor het algoritme.

2. Beleid en regelgeving

SNEL NAAR DIT ONDERDEEL

De wereldwijde ontwikkelingen op het gebied van algoritmes en AI blijven vragen om actie voor meer transparantie en verantwoordelijkheid. Daarnaast blijft beter beleid en betere regulering nodig. Dit hoofdstuk beschrijft hoe op mondiaal, Europees en nationaal niveau wordt omgegaan met de uitdagingen, kansen en risico's van AI in beleid en toezicht. Allereerst staan we stil bij de verheldering van het Hof van Justitie over transparantie en uitlegbaarheid van algoritmes. Vervolgens gaan we in op de geopolitieke uitdagingen rondom AI en de manier waarop de Verenigde Staten, China en de Europese Unie daarmee omgaan. Mede in het kader van de AI-verordening is er in Europa aandacht voor het belang van innovatie en de noodzaak van adequate bescherming van fundamentele rechten. Dit komt tot concrete uiting in de ontwikkeling van richtsnoeren, standaarden en de regulatory sandbox. De voortgang in de afgelopen maanden toont dat de EU stappen blijft zetten naar een concreter kader voor de verantwoorde inzet van algoritmes en AI in de samenleving. Het is van belang dit tempo vol te blijven houden en waar mogelijk te versnellen.

2.1 Transparantie en uitlegbaarheid van algoritmes

Wanneer de inzet van algoritmes en AI leidt tot geautomatiseerde besluitvorming, moeten mensen kunnen begrijpen hoe het systeem tot dat besluit is gekomen. Nieuwe Europese jurisprudentie verduidelijkt wanneer informatieverstrekking bij geautomatiseerde besluitvorming voldoet aan de AVG-vereisten.
Die vereisten gaan over transparantie bij het gebruik van algoritmes en AI in besluitvormende processen over personen. In februari 2025 heeft het Europees Hof van Justitie een prejudiciële beslissing genomen in een arrest dat draait om geautomatiseerde (algoritmische) beoordeling van kredietwaardigheid (Zaak C-203/22, Dun & Bradstreet Austria GmbH).[50] Het arrest gaat over een zaak waarin een telefoonabonnement met maandelijkse kosten van 10 euro werd geweigerd, op grond van ontoereikende kredietwaardigheid van de klant. In deze zaak was door een Oostenrijkse rechter eerder al geconcludeerd dat Dun & Bradstreet geen betekenisvolle informatie had verstrekt over de onderliggende logica van de kredietafwijzing. Het nieuwe arrest schept meer duidelijkheid over de omstandigheden waarin betekenisvolle informatie toereikend is.[51]

Organisaties die algoritmes en AI inzetten in hun geautomatiseerde besluitvorming moeten de uitleg niet te ingewikkeld, maar ook niet te simpel te maken. Het Hof benadrukt dat personen recht hebben op uitleg van de logica en gegevens die aan het resultaat ten grondslag liggen. Om deze informatie nuttig te maken en personen bijvoorbeeld in staat te stellen het besluit aan te vechten, moet de informatie over het besluit beknopt, transparant, begrijpelijk en gemakkelijk toegankelijk zijn.[52] Dit vereist maatwerk. Het ene uiterste is het enkel mededelen dat een algoritme is gebruikt, maar dit is onvoldoende begrijpelijk. Het andere uiterste is een gedetailleerde beschrijving van het volledige algoritme, maar dit is onvoldoende beknopt en kan in sommige gevallen niet zonder het delen van bedrijfsgeheime informatie.

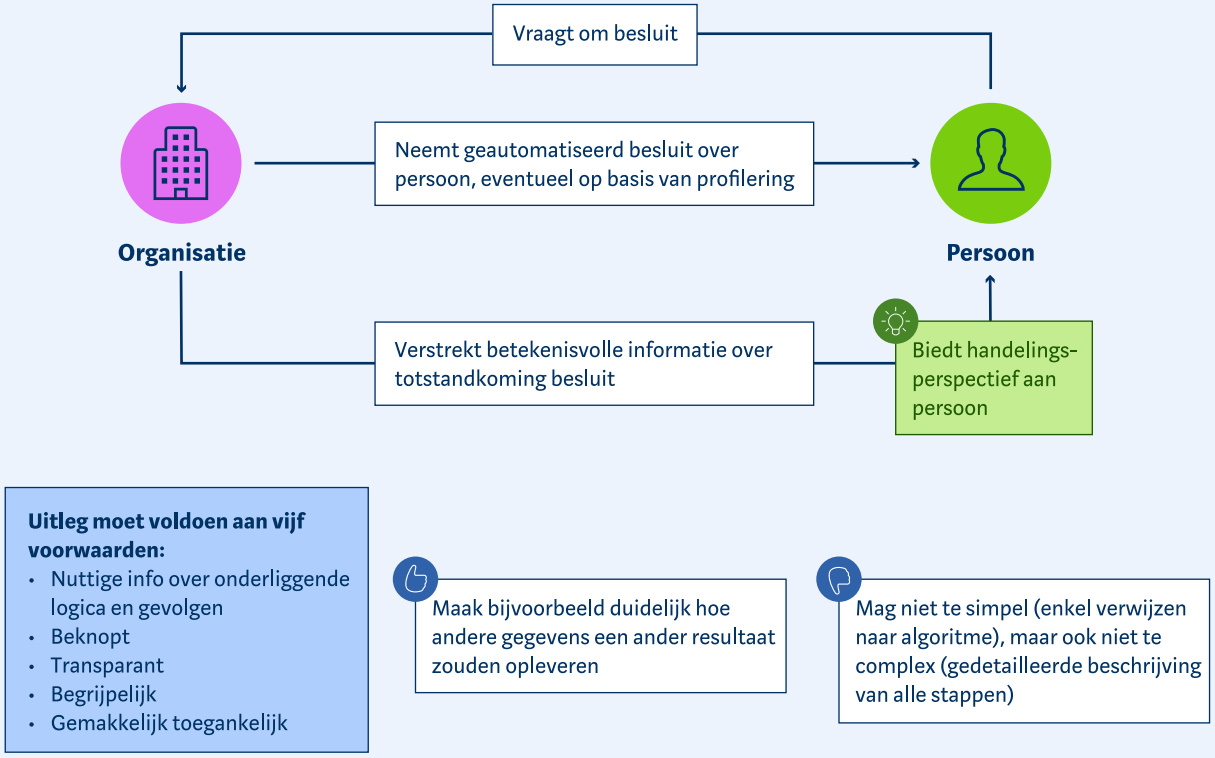
De gulden middenweg in uitlegbaarheid hangt af van de specifieke situatie. Een werkbaar voorbeeld kan zijn dat, bij geautomatiseerde afwijzing van een aanvraag, een persoon uitleg krijgt hoe specifieke aanpassingen in de gegevens de uitkomst van het besluit zouden beïnvloeden. Zo'n uitleg kan bijvoorbeeld zijn: als het bruto maandsalaris van een persoon met 500 euro stijgt, zal dezelfde kredietaanvraag met gelijke voorwaarden wel gehonoreerd worden. Tegelijkertijd geldt daarbij in de praktijk wel de uitdaging dat zo'n uitleg tijdsgebonden kan zijn (de uitleg nu biedt niet altijd een garantie voor de toekomst) en dat, bij sommige algoritmes en AI-systemen, er vele parameters kunnen zijn die een bepalende invloed hebben op de uitkomst. Een vraag waar beleidsmakers en organisaties met praktijkervaring zich dus over zullen moeten buigen is of zo'n geval inzicht moet worden gegeven in alle mogelijke manieren om goedkeuring te bereiken of dat kan worden volstaan met één mogelijke manier. Zie figuur 2.1 voor een schematische weergave.

Organisaties die algoritmes en AI inzetten in hun geautomatiseerde besluitvorming, moeten voldoende

aandacht besteden aan hun uitleg. Immers, om betekenisvolle uitleg te geven over de besluitvorming van het systeem, moet het besluit logisch te herleiden zijn uit de werking van het algoritme. Dit kan bij al te complexe algoritmes lastig zijn. Maar, wanneer deze uitleg ontbreekt, dan belemmert dit ook het recht op bezwaar of een mogelijke correctie van het besluit. Het is dan immers onvoldoende duidelijk wat iemand moet veranderen om de organisatie tot een ander besluit te laten komen. Bovendien kan de persoon niet controleren of het besluit wel zorgvuldig is gemaakt. De consequentie van dit uitlegbaarheidsvereiste is dus dat een ontwikkelaar van een algoritme of AI-systeem vanaf het begin af aan rekening moet houden met de noodzaak van menselijke autonomie. Personen over wie een algoritme of AI een geautomatiseerd besluit neemt, moeten daar uitleg over krijgen.

FIGUUR 2.1 | INTERACTIEPROFILERING, BESLUITVORMING EN INFORMATIEVERSTREKKING BIJ ALGORITMES

Betekenisvolle informatieverstrekking over algoritmes biedt handelingsperspectief aan personen bij besluiten door overheid en bedrijven.



Deze uitleg geeft ook nadere inkleuring in het licht van de AI-verordening. De prejudiciële beslissing van het Hof verduidelijkt de toepassing van het inzagerecht bij geautomatiseerde besluitvorming, zoals vastgelegd in de AVG.[53] Daarnaast geeft de AI-verordening het recht op uitleg bij individuele besluitvorming.[54] Kort gezegd, houdt dit recht in dat personen die getroffen worden door een besluit van een hoogrisico-AI-systeem, duidelijke en relevante informatie moeten ontvangen. De uitleg moet gaan over de rol van het AI-systeem in de besluitvorming en de voornaamste elementen van het genomen besluit. Dit raakt sterk aan de soortgelijke bepaling die volgt uit de AVG. De beslissing van het Hof in de zaak Dun & Bradstreet geeft daarmee ook inkleuring aan de wijze waarop AI-ontwikkelaars en AI-gebruikers hun systemen moeten ontwerpen en documenteren om aan de uitlegbaarheids-eisen te voldoen bij individuele besluitvorming.[55] Vanuit het gegevensbeschermingstoezicht werkt de AP dit jaar aan verdere uitleg over betekenisvolle informatieverstrekking.

2.2 Internationale ontwikkelingen

Op internationaal vlak gaat de aandacht naar de geopolitieke ontwikkelingen rondom AI. In de globale AI-race verschuiven de machtsverhoudingen. Voor-namelijk de Verenigde Staten en China concurreren om een leidende rol in de ontwikkeling van een infrastructuur

rondom AI. De introductie van Deepseek - dat zowel als chatdienst als open source model beschikbaar is gekomen - markeert de toegenomen rol van China in de wereldwijde AI-infrastructuur. De snelle ontwikkelingen op het gebied van AI gaan gepaard met toenemende geopolitieke vraagstukken. Dit komt mede door de ingrijpende impact van deze technologie op mondiale machtsverhoudingen, economische structuren en maatschappelijke processen.

De VS en China hebben hun leidende rol te danken aan forse investeringen in AI. Het AI Index Report 2025 van het *Stanford Institute for Human-Centered AI* concludeert onder andere dat de kloof in de race om leiderschap tussen de VS en China aanzienlijk verminderd is. Waar de VS nog altijd vooroploopt in private investeringen en het produceren van AI-modellen, heeft China de kloof gedicht met kwaliteitsverbeteringen. Ook blijft China leider op het gebied van AI-publicaties en uitgegeven patenten.[56]

De VS legt sinds begin dit jaar publiekelijk de aandacht op een nieuwe koers van de simplificatie en deregulering van AI. Tegelijkertijd werkt de VS ook aan concrete regelgeving die gelijkenissen vertoont met de AI-verordening.[57] Bijvoorbeeld via een uitvoeringsbevel met als doel barrières voor AI-innovatie weg te nemen en de leiderschapspositie van de VS te behouden.[58] Dit uitvoeringsbevel vervangt het eerdere uitvoeringsbevel van de regering-Biden, waarin nadruk lag op de randvoorwaarden voor de inzet van veilige en betrouwbare AI binnen de federale overheid. Ondanks de nadruk op simplificatie van regulering en de zorgen over veiligheid en bescherming van individuele rechten, bevat een recent aangekondigd memorandum gelijksoortige waarborgingsbepalingen als in het eerdere uitvoeringsbevel. De memo *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust* kondigt verplichtingen aan voor 'high-impact AI', waaronder documentatie, impact assessments, testen en mogelijkheden tot menselijke tussenkomst.[59] Opvallend is dat de definitie van 'high impact AI' gelijkenissen vertoont met de AI-verordening. Zo wordt AI dat dient als basis voor beslissingen met gevolgen voor toegang tot onderwijs, wonen, verzekeringen of arbeid als hoog risico aangemerkt.[60]

China werkt aan een reguleringskader met aandacht voor concretere standaarden. Zo moet vanaf 1 september 2025 AI-gegenereerde content duidelijk herkenbaar zijn voor gebruikers. Er wordt gewerkt aan nationale AI-standaarden en China neemt actief deel aan de ontwikkeling van internationale standaarden.[61]

Wereldwijd zijn de private investeringen in AI flink gestegen, een toename van 44.5% tussen 2023 en 2024. Buiten de VS, de EU-lidstaten en China zijn ook andere landen fors aan het investeren. In 2024 heeft het Verenigd Koninkrijk 4.52 miljard US dollar geïnvesteerd, Canada 2.89 miljard en de Verenigde Arabische Emiraten 1.77 miljard.[62] Dit laat zien dat veel landen de kansen van AI inzien.

Box 2.1

Wetgeving als randvoorwaarde voor innovatie

Hoewel regulering en innovatie vaak als tegenpolen worden gepresenteerd, zijn er ook sterke argumenten om wetgeving juist te zien als aanjager van verantwoorde innovatie.

- **Ten eerste** creëert Europese wetgeving gelijke spelregels en rechtszekerheid. Bedrijven weten waar ze aan toe zijn, ongeacht in welke lidstaat ze opereren. Dit voorkomt juridische versnippering en maakt het aantrekkelijker om innovatieve producten op Europese schaal uit te rollen.
- **Ten tweede** voorkomt wetgeving dat bedrijven concurreren op aspecten waar minimumnormen gelden, zoals privacy of veiligheid. Hierdoor kunnen zij zich juist richten op de innovatie van gebieden waar dat wél gewenst is. Wetgeving stuurt innovatie daarmee in maatschappelijk verantwoorde en toekomstbestendige richtingen.
- **Ten derde** vergroot wetgeving het vertrouwen van burgers in digitale diensten. Dat vertrouwen ondersteunt de brede acceptatie en het gebruik van nieuwe technologieën.

Tot slot biedt Europese wetgeving steeds vaker ruimte voor innovatie. Goede voorbeelden zijn de verplichte *regulatory sandboxes* in de AI-verordening. Door samenwerking tussen toezichthouders en ontwikkelaars mogelijk te maken, kunnen ontwikkelaars verantwoord innoveren.

De AP ziet ook in Europa een duidelijke beweging naar simplificatie van de regelgeving. Het juridische raamwerk in Europa legt momenteel een sterke nadruk op fundamentele rechten en publieke waarden, zoals transparantie. Tegelijkertijd ontstaat ook in Europa bezorgdheid over het internationele vermogen op het gebied van AI en zijn er vragen over de complexiteit van het juridische raamwerk en de gevolgen voor Europa's concurrentiepositie. De aandacht hiervoor groeide nog sterker na het Draghi rapport, dat verscheen in september 2024. In dit rapport werd gewaarschuwd dat de complexe regelgeving de positie van Europa benadeelt ten opzichte van de VS en China.[63] Onlangs ondertekenden dertien Europese lidstaten een verklaring waarin ze oproepen tot meer investeringen in AI, het wegnemen van belemmeringen en het vereenvoudigen van EU-regels en -procedures.[64] Te midden van deze discussies heeft de Europese Commissie de afgelopen tijd een aantal maatregelen aangekondigd. Een daarvan is het *AI Continent Action Plan*, met als doel Europa wereldleider te maken op het gebied van AI.

Het AI Continent Action Plan gaat vooral over het bevorderen van AI-innovatie in de EU. Zo moet de Europese AI-infrastructuur opgeschaald worden om onderzoek en innovatie mogelijk te maken voor het trainen en finetunen van AI-modellen. Ook wordt geïnvesteerd in het waarborgen van hoge kwaliteit data voor AI-innovaties. Verder zet het plan in op het vergroten van AI-vaardigheden en AI-literacy.[65]

De Europese bereidheid om te investeren in AI en algoritmes is groot. Dit blijkt uit het hiervoor beschreven Europese AI Continent Action Plan. Daarnaast lanceerde de EU onlangs InvestAI, een initiatief van de EU om 200 miljard euro aan investeringen voor AI vrij te maken. Een fonds van 20 miljard zal toekomstige AI-gigafabrieken financieren, waar complexe, zeer grote AI-modellen worden getraind. In Nederland heeft het kabinet aangekondigd te investeren in een AI-fabriek die ook onderdeel wordt van het Europese netwerk van AI-faciliteiten van EuroHPC.[66] De AI-fabriek moet een one-stop-shop worden voor AI-ontwikkeling door als kenniscentrum te functioneren, AI-rekenkracht te bieden en dataopslag te faciliteren.

De roep om vereenvoudiging van regelgeving vindt weerklank binnen de EU. Maar er klinken ook andere geluiden... Zo benadrukken experts dat het Europese digitale regelgevingskader juist is ontworpen om innovatie te bevorderen en het vertrouwen te vergroten door potentiële risico's te beheersen. Anderen wijzen erop dat investeringen in Europa juist achterblijven door het gebrek aan private investeringen. Dit komt bijvoorbeeld door zwaktes in de kapitaalmarktunie en tekortkomingen in regels over overheidssteun.[67]

Box 2.2

Platformwerkers beter beschermd

De Europese Platformwerkrichtlijn is een belangrijke stap in de bescherming van platformwerkers die door algoritmes en AI in hun werk gestuurd worden. Denk bijvoorbeeld aan mensen die via platformen eten bezorgen of anderen vervoeren. De Platformwerkrichtlijn is eind 2024 aangenomen en moet de komende tijd omgezet worden in Nederlandse regelgeving. De richtlijn verduidelijkt de (rechts)positie van platformwerkers en streeft naar verbetering van arbeidsomstandigheden. Zo komen er aanvullende regels voor monitoring en algoritmisch management van platformwerkers. Het gaat dan om het waarborgen van taaktoedeling en de beoordeling van platformwerkers, ook op het gebied van hun veiligheid.
Verder heeft de richtlijn als doel de transparantie rondom algoritmisch management te verbeteren. Zo moeten ondernemingsraden en vakbewegingen bijvoorbeeld ook worden geïnformeerd.
De regels in de richtlijn hangen sterk samen met andere (digitale) regelgeving, zoals de AVG, de AI-verordening en de DSA. Maar de regels houden ook verband met regelgeving over een gezonde en veilige werkomgeving. De richtlijn is een voorbeeld van regelgeving die op sterk veralgoritmiseerd terrein aanvullende en specifieke waarborgen biedt.

2.3 Nationale ontwikkelingen

Op nationaal niveau valt het laatste half jaar vooral het werk van de staatssecretaris Digitalisering op, net als enkele aangenomen moties in de Tweede Kamer. Ook gaat het toezicht op de Digitale Dienstenverordening (DSA) van start.

Zo publiceerde het kabinet in april het nieuwe standpunt over de inzet van generatieve AI bij de overheid.[68] Het nieuwe standpunt is minder voorzichtig en moedigt het gebruik van generatieve AI juist aan. Het standpunt gaat samen met een handreiking die overheidsorganisaties moet helpen generatieve AI verantwoord in te zetten. Ook moeten zij afspraken maken met leveranciers en gelden er aparte voorwaarden. De AP verwacht dat overheden generatieve AI door dit standpunt meer gaan gebruiken. Alhoewel de AP vindt dat zorgvuldigheid geboden is, ziet het ook dat in het standpunt en de handreiking veel goede ingrediënten zijn opgenomen voor de inzet van generatieve AI.

De AP benadrukt echter dat een aantal randvoorwaarden voor verantwoorde inzet op toepassingsniveau meer aandacht verdienen. Dit ziet voornamelijk toe op de randvoorwaarden voor gebruik van generatieve AI in de praktijk. Een voorbeeld hiervan is het borgen van voldoende AI-geletterdheid. Ook is het van groot belang dat generatieve AI in bredere context wordt gezien. Wanneer dit niet gebeurt, dreigt de inzet van generatieve AI bestaande processen (onbedoeld) ingrijpend te veranderen. In onze rapportage uit het najaar van 2023 schreven we meer over zogenoemde algoritmevervorming.[69]

De AP draagt zelf ook bij aan het maatschappelijk debat om de verantwoorde inzet van generatieve AI te stimuleren. Hiervoor publiceerde de AP op 23 mei een consultatieversie van een eigen visie op generatieve AI. De AP ging ook over de visie in gesprek met partijen die generatieve AI ontwikkelen, gebruiken en onderzoeken. Naast het faciliteren van het debat, schetst de AP hierin de contouren van een gewenst toekomstperspectief voor een rechtmatige en verantwoorde inzet van generatieve AI.

Verder valt op dat de afronding van de internetconsultatie 'Algoritmische besluitvorming en de Awb' verder is uitgesteld.[70] Deze consultatie liep tot april 2024 en de AP heeft in een eerdere rapportage al aangegeven uit te kijken naar het vervolg. Juist bij de verdere digitalisering van de overheid, is het belangrijk het (juridische) systeem eromheen te betrekken. Zo blijft het handelen van de overheid controleerbaar, uitlegbaar en rechtvaardig.

Tot slot heeft de regering de afgelopen tijd verder gewerkt aan de ontwikkeling van een wetenschappelijke standaard voor het gebruik van modellen en algoritmes.[71] Op basis van werk van de Universiteit Leiden wordt deze standaard verder ontwikkeld. De AP benadrukt nogmaals dat de eisen hiervoor in samenhang met de bepalingen uit de AI-verordening bekeken moeten worden. Het is onwenselijk dat vergelijkbare eisen in verschillende (product)standaarden anders uitgelegd kunnen worden.

De Tweede Kamer nam op 20 mei een motie aan die de regering oproept meer werk te maken van algoritmeregistratie.[72] De motie ziet specifiek toe op de algoritmes die mogelijk gebruikmaken van risicoprofilering en geautomatiseerde selectie-instrumenten. Een eerder verzoek om dit soort algoritmes te publiceren, is nog niet volledig uitgevoerd. Deze motie probeert daar terecht verandering in te brengen.

Box 2.3

Nederlands toezicht op de DSA van kracht

De DSA is sinds 17 februari 2024 volledig van kracht. De DSA waarborgt de bescherming van grondrechten in de onlinewereld. De wet geldt voor aanbieders van "tussenhandeldiensten", zoals hostingbedrijven, sociale media, onlinemarktplaatsen, accommodatieplatforms, onlineplatforms en zoekmachines. De regels moeten transparantie en de rechten van gebruikers waarborgen. Ook moet de wet online misleiding en illegale informatie tegengaan.

Het Europese toezicht op de allergrootste platforms was al van start. Zo stelde de Europese Commissie in voorlopige bevindingen van een onderzoek al eerder dat TikTok de DSA overtreedt.[73] Het bedrijf zou niet voldoen aan de verplichting om een archief van advertenties bij te houden. TikTok krijg nu de kans om te reageren, voordat een definitief besluit wordt genomen. Daarnaast stelt de Commissie momenteel richtsnoeren op voor de bescherming van kinderen. Een eerste concept voor een publieke consultatie is gepubliceerd.[74] De richtsnoeren bevatten een niet-uitputtende lijst met maatregelen voor platforms om minderjarigen te beschermen en om te voldoen aan de DSA.

In Nederland houden de Autoriteit Consument en Markt (ACM) en de AP per 4 februari 2025 toezicht op de DSA. In de Uitvoeringswet is bepaald dat de ACM is aangewezen als digitale dienstencoördinator en toezichthouder op een groot deel van de DSA. De AP houdt toezicht op de regels over persoonsgegevens. Daarnaast houdt de AP nog toezicht op de regels over transparante aanbevelingsalgoritmes. Deze toezichtstaak is bij de AP belegd om een breed perspectief van burgers op aanbevelingsalgoritmes te waarborgen.[75]

2.4 Verduidelijking en concretisering AI-verordening

Ruim een jaar na de start van de eerste onderdelen van de AI-verordening worden ook stappen zichtbaar voor de verdere verduidelijking en concretisering van de wet. Zowel de AI-Office als de beoogde nationale toezichthouders werken aan initiatieven om de naleving van de verordening te faciliteren. Dit gebeurt onder andere via richtsnoeren, de praktijkcode voor general purpose AI en gedragscodes.

Richtsnoeren

De Europese Commissie publiceerde in 2025 de eerste richtsnoeren over de AI-verordening. De eerste twee richtsnoeren over (1) de verboden in de AI-verordening en (2) de definitie van AI-systemen zijn in februari gepubliceerd. Richtsnoeren over andere onderwerpen zullen de komende jaren volgen.

De AP ziet dat er ook met de richtsnoeren veel ruimte blijft voor verdere concretisering en kaderstelling. De richtsnoeren worden gepubliceerd als een '*living document*', dat de Europese Commissie up-to-date zal houden en kan aanpassen als dat nodig is. Dit gebeurt bijvoorbeeld op basis van de ervaring van toezichthouders, nieuwe jurisprudentie of technologische ontwikkelingen.

De Europese Commissie geeft in de richtsnoeren over verboden AI een brede uitleg aan de verantwoordelijkheid van AI-aanbieders om verboden praktijken te voorkomen.[76] AI-aanbieders moeten zelf goed in kunnen schatten in hoeverre hun AI-systeem wordt gebruikt door de afnemers van hun product. Op basis van deze inschatting moeten zij waarborgen inbouwen om verboden gebruik te voorkomen. Zo moet een aanbieder van een emotieherkenningssysteem er bijvoorbeeld voor waken dat deze niet wordt gebruikt op scholen of op de werkplek. Daarnaast geeft de Europese Commissie in de uitgebreide richtsnoeren een concretere invulling aan de verboden, onder andere met voorbeelden.

De AP ziet dat er ondanks de richtsnoeren nog veel onduidelijkheid blijft bestaan over de definitie van het begrip "AI-systeem".[77] Zo wordt onvoldoende duidelijk of en hoe een systeem kan bestaan uit meerdere onderdelen of processen. Ook is niet duidelijk of de interface van een systeem bijvoorbeeld ook tot het 'AI-systeem' behoort. Onduidelijkheid ontstaat ook over het vereiste 'inferentievermogen' en voorbeelden van systemen die buiten de definitie vallen. Positief is dat de Europese Commissie in de richtsnoeren wel benadrukt dat de AI-systeemdefinitie gaat over zowel de ontwikkel-, als de gebruiksfase van het systeem. Daarmee valt ook het gebruik van AI-technologie voor eenvoudigere algoritmes binnen de reikwijdte van de AI-verordening.

De komende tijd zullen meer richtsnoeren van de Europese Commissie volgen. De AP verwacht deze zomer in ieder geval aanvullende richtsnoeren over de meldplicht voor ernstige incidenten bij hoogrisico-AI-systemen. Daarnaast stelt de AI-verordening dat er voor 2 februari 2026 richtsnoeren beschikbaar moeten zijn over de classificatieregels voor AI-systemen met een hoog risico. Hiermee moet niet alleen duidelijker worden welke soort AI-systemen in deze categorie vallen, maar ook in welke gevallen een uitzonderingsbepaling geldt. Zie figuur 2.2.

FIGUUR 2.2 | TIJDLIJN PUBLICATIE RICHTSNOEREN

Code of Practice on GPAI

Een zeer impactvolle concretisering van de AI-verordening is de praktijkcode voor *general purpose AI*. Naleving van de praktijkcode is voor aanbieders van *general purpose AI* een manier om aan te tonen dat ze aan de wet voldoen. In de vorige risicorapportage schreven we hier uitgebreider over.[78] Tijdens het schrijven van deze rapportage is de definitieve versie van de praktijkcode helaas nog niet gepubliceerd. Omdat AI-modellen voor algemene doeleinden vanaf augustus wel al aan de strengere regels uit de verordening moeten gaan voldoen, wordt de praktijkcode echter ook snel verwacht.

Standaarden

Europese standaarden voor de praktische uitwerking van de vereisten aan hoogrisico-AI-systemen laten voorlopig op zich wachten. De deadline in het eerste standaardisatieverzoek (1 april 2025) van de Commissie aan de standaardisatieorganisatie JTC-21 is inmiddels verlopen. Bovendien is het nog niet duidelijk wanneer alle standaarden dan wel zullen worden opgeleverd. Gezien de huidige voortgang raadt de AP AI-aanbieders aan om alvast te beginnen met de naleving van de verordening, op basis van de wettekst en waar mogelijk met good practices. Het is zaak niet op deze standaarden te blijven wachten. Ondertussen is het aan beleidsmakers en toezichthouders om, nog voordat de vereisten voor de eerste AI-systemen met hoog risico ingaan in augustus 2026, helder te maken wat dit voor de naleving van de AI-verordening en het toezicht daarop betekent.

Overige initiatieven

Naast richtsnoeren, gedragscodes, sjablonen en standaarden zijn er meer initiatieven ter verduidelijking van de AI-verordening. Zo startte de Commissie aan een AI Pact dat AI-aanbieders moet helpen bij de implementatie van de verordening. Er zijn de afgelopen maanden ook verschillende webinars georganiseerd die voor iedereen toegankelijk zijn. De webinars geven deelnemers een beter inzicht in de verordening en de implementatie ervan. De webinars zijn terug te kijken.[79]

De AI-Office werkt aan het opstarten van een AI-verordening Service Desk. Het doel van de service-desk is om bedrijven en overheden te ondersteunen bij de naleving van de verordening. Tot 19 mei stond een aanbesteding[80] open voor een extern team dat, onder aansturing van de AI-Office, advies gaat geven over vragen van AI-aanbieders over de wet. De servicedesk wordt een interactief platform dat, als centraal knooppunt, ook

algemene informatie en ondersteuningsmateriaal verstrekt. Het is nog niet bekend wanneer de servicedesk start.

Verder wordt er zowel op nationaal als Europees niveau gewerkt aan invulling van de verplichting tot AI-geletterdheid. De AI-Office werkt hiervoor aan een gedragscode binnen de kaders van het AI Pact. Daarnaast heeft de Europese Commissie recent een FAQ gepubliceerd. Op nationaal niveau heeft de AP een oproep tot input[81] geplaatst om praktijkvoorbeelden van AI-geletterdheid te verzamelen. Op basis van de antwoorden stelt de AP een document op met goede praktijkvoorbeelden. Nederlandse organisaties kunnen hier inspiratie uithalen voor hun eigen AI-geletterdheidsbeleid. De resultaten op de uitvraag werden gepresenteerd in het seminar van de AP op 18 juni over AI-geletterdheid.

Tot slot heet de Europese *public buyers community* een bijgewerkte versie van de al bestaande model-contractbepalingen gepubliceerd.[82] Deze zijn opgesteld om inkopers van AI-systemen in de publieke sector te ondersteunen. De update bevat een modelcontract voor de inkoop van AI met een hoog risico en is volledig afgestemd op de definitieve versie van de AI-verordening. Bovendien zijn bepalingen aanpasbaar naar specifieke behoeften bij de inkoop van AI zónder hoog risico. Ook is er een toelichting voor het gebruik van de bepalingen. Modelcontracten zijn nuttig, omdat het bij de aanschaf van een AI-systeem belangrijk is de inkoopvoorwaarden op orde te hebben. Zo moet je bijvoorbeeld ook goed weten wat er gebeurt met de data die het AI-systeem verwerkt en hoe het zit met de regeling van de rechten op die data.

2.5 Nationale ontwikkelingen AI-verordening

In maart publiceerde de AP samen met de Rijksinspectie voor Digitale Infrastructuur een voorstel voor de Nederlandse invulling van de *regulatory sandbox*.[83] Dit voorstel is opgesteld in samenwerking met diverse toezichthouders en ministeries, en beschrijft de gewenste uitgangspunten en de procesinrichting van de sandbox. Het doel van het voorstel is om een stevige basis te leggen voor de verdere samenwerking en afstemming. Uiterlijk in augustus 2026 start de definitieve sandbox. Voor het vormvoorstel is een pilot doorlopen met vragen van verschillende organisaties over de AI-verordening. Ook zijn er gesprekken gevoerd met stakeholders uit het Nederlandse AI-ecosysteem.

Het vormvoorstel beschrijft hoe de Nederlandse sandbox zo optimaal mogelijk kan bijdragen aan rechtszekerheid en innovatie. Daartoe is het belangrijk dat alle relevante toezichthouders onder de AI-verordening actief deelnemen. Zo kan toegang tot de sandbox via één centraal loket worden geregeld. Dit voorkomt dat AI-aanbieders zelf moeten achterhalen met welke toezichthouder zij contact moeten opnemen. Bovendien draagt deze werkwijze bij aan een consistente uitleg van wet- en regelgeving door de verschillende toezicht-houders. Tot slot maakt deze werkwijze het mogelijk om verwante regelgeving in samenhang te duiden.

De meerwaarde van de sandbox ligt in het onder-steunen van AI-aanbieders bij het voldoen aan wet- en regelgeving. Dit doen zij bijvoorbeeld via juridisch en technisch advies over de interpretatie, het testen of valideren van de wet. Omdat toezichthouders niet in alle behoeften kunnen voorzien, moet de sandbox goed aansluiten op andere initiatieven, zoals AI-factories, Testing and Experimentation Facilities en EDIH's.

Naar aanleiding van bovenstaand voorstel is op 20 mei een Tweede Kamermotie aangenomen om de regulatory sandbox met prioriteit uit te werken. De motie roept de regering op ervoor te zorgen dat deze uiterlijk in het eerste kwartaal van 2026 beschikbaar is. De AP verwelkomt deze oproep en zet zich ervoor in om zo snel mogelijk met een sandbox te starten. De AP werkt samen met de departe-menten om dit ook mogelijk te maken.

Box 2.4

AI, privacy en datagovernance: De OECD-aanpak van betrouwbare AI

Door: The OECD Directorate for Science, Technology and Innovation (STI)

AI verandert snel de manier waarop samenlevingen gegevens gebruiken en delen en ontsluit nieuwe kansen voor innovatie. Maar het roept ook kritische vragen op over privacy, vertrouwen en verantwoord databeheer. De OECD (Organisation for Economic Co-operation and Development) is voorloper bij het vormgeven van internationale beleidskaders die ervoor moeten zorgen dat mensenrechten en democratische waarden gevolgd worden bij de ontwikkeling en toepassing van AI. In 2019 heeft de OECD 's werelds eerste intergouvernementele norm voor AI vastgesteld – de AI-beginselen van de OECD – die in mei 2024 zijn geactualiseerd. De AI-beginselen zijn aangenomen door meer dan 47 landen, waaronder grote economieën en internationale partners. Ze bieden een blauwdruk voor betrouwbare, mensgerichte AI en ondersteunen nationale en internationale beleidskaders.

Beleidsinstrumenten ter bevordering van betrouwbare AI

Het OECD.AI Policy Observatory ondersteunt de implementatie van de AI-beginselen. Het biedt empirisch onderbouwde inzichten, realtime gegevens en beleids-instrumenten om de transparantie te vergroten, internationale samenwerking te vergemakkelijken en verantwoorde AI-innovatie te ondersteunen.

Belangrijke instrumenten zijn bijvoorbeeld de Global AI Initiatives Navigator (GAIIN), een live-repository van ruim 1.300 AI-beleidsinitiatieven uit meer dan 70 landen en intergouvernementele organisaties met interactieve visualisaties van wereldwijde AI-trends. Het Policy Observatory omvat praktische tools zoals de AI-monitor voor incidenten (AIM) en de OECD-catalogus van instrumenten en maatstaven voor betrouwbare AI, die een gecureerd register van instrumenten en maatstaven biedt om ontwikkelaars en aanbieders te ondersteunen bij de integratie van robuuste en betrouwbare procedures tijdens de levenscyclus van het AI-systeem. De AI Wonk-blog draagt verder bij aan dit werk door experts een platform te bieden om inzichten te delen over hoe betrouwbaar AI-beleid het best kan worden vormgegeven.

Betrokkenheid van experts bij prioriteiten

De expertgemeenschap van OECD.AI en Global Partnership on AI (GPAI) bestaat uit experts van de overheid, het bedrijfsleven, de academische wereld en het maatschappelijk middenveld. De gemeenschap ondersteunt de implementatie van de AI-beginselen van de OECD en informeert het Policy Observatory via specifieke expertgroepen. Deze expertgroepen bepalen belangrijke prioriteiten en bieden een platform om kansen en uitdagingen op het gebied van AI-beleid te bespreken.

De expertgroepen voor AI-incidenten en voor AI, data & privacy richten zich op belangrijke kwesties zoals inzicht in de risico's die verbonden zijn aan de ontwikkeling en adoptie van AI-systemen. En ze onderzoeken hoe privacy en gegevensbescherming bijdragen aan betrouwbare AI.

Definiëren, rapporteren en monitoren van AI-incidenten

Met de snelle ontwikkelingen en adoptie van AI-systemen beginnen de bijbehorende risico's zich al te manifesteren. Het begrijpen van deze gebeurtenissen en hun implicaties is een prioriteit geworden voor beleidsmakers. In februari 2025 heeft de OECD een gemeenschappelijk rapportagekader voor AI-incidenten gepubliceerd, dat een flexibele structuur biedt voor het melden en monitoren van AI-incidenten. Het kader is gezamenlijk ontwikkeld met de expertgroep voor AI-incidenten. Het bevat 29 criteria om een AI-incident te beschrijven en te melden, deels op basis van het OECD-kader voor de classificatie van AI-systemen. Het rapporteren en monitoren van AI-incidenten zal beleidsmakers, ontwikkelaars, gebruikers en andere belanghebbenden in staat stellen risicovolle systemen in verschillende contexten te identificeren, inzicht te krijgen in zowel huidige als ontwikkelende risico's en hun impact op getroffen personen, groepen, rechten of waarden te evalueren.

In dit verband heeft de OECD in november 2023 de AI Incidents and Hazards Monitor (AIM) gelanceerd.

De AIM documenteert AI-incidenten en -gevaren om belanghebbenden te helpen waardevolle inzichten te krijgen in de risico's en impact van AI-systemen. De AIM zal de identificatie van AI-risicopatronen, ook die met betrekking tot privacy en data governance, op zowel mondiaal als regionaal niveau vergemakkelijken. En het inzicht in de veelzijdige aard van impact en schade als gevolg van de ontwikkeling, het gebruik en het functioneren van AI-systemen vergroten.

Evenwicht tussen innovatie, privacy en data governance

AI is afhankelijk van enorme datasets om effectief te functioneren. Soms bevatten die sets opzettelijk of onbedoeld persoonsgegevens. Dit leidt tot aanzienlijke privacyrisico's, waaronder mogelijk misbruik van persoonsgegevens, bias en een gebrek aan transparantie over de manier waarop persoonsgegevens worden verzameld, verwerkt en gedeeld. De privacyrichtlijnen van de OECD bieden sterke basisbeginselen voor de bescherming van privacy en gegevensrechten in het tijdperk van AI. Deze richtlijnen worden sinds 1980 als wereldwijde minimumnorm erkend en zijn in 2013 bijgewerkt

In 2024 heeft de OECD een deskundigengroep inzake AI, data & privacy opgericht om synergieën te verkennen en gecoördineerde beleidsoplossingen te ontwikkelen in AI- en privacygemeenschappen. Dit initiatief is gebaseerd op de erkenning dat deze gemeenschappen van oudsher afzonderlijk opereren, wat kan leiden tot versnipperde benaderingen en lacunes in de regelgeving. Een belangrijk resultaat van deze groep was het in kaart brengen hoe OECD-privacyrichtsnoeren zich verhouden tot de AI-beginselen van de OECD, die bedoeld zijn om beleidsmakers en beroepsbeoefenaars te helpen de kansen van AI-gedreven innovatie in evenwicht te brengen, met de noodzaak om overwegingen op het gebied van privacy en datagovernance te integreren in de hele AI-levenscyclus, van ontwerp tot uitrol.

Vanuit deze basis onderzoekt de OECD ook de gevolgen voor privacy en gegevensbescherming van de mechanismen die ingezet worden voor het verzamelen van gegevens voor datasets voor het trainen van AI-modellen. Dit omvat praktijken zoals data scraping, die complexe problemen opwerpen op het snijvlak van intellectueel eigendom en privacy-rechten. Het OECD-rapport over intellectuele eigendomskwesties in AI die zijn getraind op basis van gescrapte data biedt een overzicht van de wijze waarop data scraping voor AI-ontwikkeling interacteert met verschillende intellectuele-eigendomsrechten. De OECD zal in de toekomst ook de gevolgen voor de privacy en de gegevensbescherming onderzoeken van andere relevante mechanismen voor gegevensverzameling die worden gebruikt voor het bouwen van datasets voor AI-training.

Toegang tot hoogwaardige gegevens is van cruciaal belang voor de ontwikkeling van AI-modellen, maar de toegang tot deze gegevens moet in balans zijn met robuuste bescherming. De aanbeveling van de OECD over het verbeteren van de toegang tot en het delen van gegevens biedt een kader voor het balanceren van open toegankelijkheid van gegevens met legitieme bescherming en waarborgen, waarbij het beginsel wordt toegepast om gegevens "zo open als mogelijk, zo gesloten als nodig" te maken. In het OECD-verslag over het verbeteren van de toegang tot en het delen van gegevens in het tijdperk van AI wordt benadrukt hoe overheden de toegang tot en het delen van gegevens en bepaalde AI-modellen kunnen verbeteren en tegelijkertijd de privacy en andere rechten en belangen kunnen waarborgen.

Als aanvulling op deze werkzaamheden bevordert de OECD actief privacybevorderende technologieën (PET's) als een belangrijk onderdeel van verantwoorde AI- en data governance strategieën. Wanneer PET's op de juiste manier worden gebruikt, zoals *federated learning*, *homomorphic encryption*, of *differential privacy*, kunnen AI-modellen worden getraind en ingezet met verminderd gebruik van - en verminderd risico voor persoonsgegevens. Zo wordt privacy by design gedurende de hele AI-levenscyclus ondersteund.

Samen ondersteunen deze initiatieven een evenwichtige aanpak van datagovernance, waarbij innovatie mogelijk wordt gemaakt terwijl individuele rechten worden behouden en het vertrouwen in AI wordt versterkt.

Bezoek voor meer informatie:
https://oecd.ai
https://oecd.ai/site/incidents
https://oecd.ai/site/data-privacy

3. AI en emotieherkenning

SNEL NAAR DIT ONDERDEEL

Emoties zijn een essentieel onderdeel van het menselijk bestaan. Ze spelen een bepalende rol in ons dagelijks leven: van sociale interactie, tot onze waarneming, hoe we leren, welke beslissingen we nemen en hoe. Er zijn steeds meer AI-systemen waarvan geclaimd wordt dat ze emoties kunnen herkennen op basis van biometrie. De markt voor toepassingen van zulke systemen heeft de afgelopen jaren een gestage groei doorgemaakt.[84] [85] Toch zijn de fundamentele aannames onder de systemen wankel. De werking is daardoor twijfelachtig. Als de systemen toch worden ingezet, ontstaan er bovendien risico's op inbreuk van grondrechten en publieke waarden. Zo kan de inzet leiden tot discriminatie, inperking van menselijke autonomie en privacyschendingen.

3.1 Emotieherkenning op basis van biometrie

Biometrische gegevens worden steeds meer ingezet om emoties van mensen te herkennen. Biometrische gegevens zijn lichaams- en gedragskenmerken van mensen. Deze worden al lange tijd gebruikt voor verificatie en identificatie van personen. Automatische analyse van biometrie en gebruik van meer databronnen heeft de ontwikkeling van nieuwe toepassingen mogelijk gemaakt.[86] Er zijn nu dus ook steeds meer systemen die als doel hebben om emoties te herkennen. Deze systemen analyseren biometrische gegevens en proberen emoties te herkennen. Dit kan de basis vormen voor beslissingen, aanbevelingen of andere output. Biometrische systemen kijken dus niet alleen meer naar *wie je bent*, maar ook naar *hoe het met je gaat*.[87]

AI in de jaren '70 legden onderzoekers, die emoties probeerden te identificeren op basis van fysieke en fysiologische signalen, de basis voor emotie-herkenningssystemen.[88] Met de opkomst van computers ontwikkelden onderzoekers algoritmes die gezichts-uitdrukkingen en stemgeluiden konden analyseren. De doorbraak kwam met de opkomst van *machine learning*, waardoor AI grote hoeveelheden data kon verwerken. Hierdoor werden de systemen toegankelijker en nauwkeuriger. Sindsdien zien bedrijven en andere organisaties de kans om deze techniek in te zetten voor verschillende toepassingen.[89]

Verschillende sectoren kijken tegenwoordig naar diverse emotieherkenningssystemen, of gebruiken deze al. Er zijn systemen die via gezichtsuitdrukkingen proberen te bepalen welke emoties iemand ervaart. Ook zijn er systemen die huidgeleiding, hartslag of stemgeluid meten en beoordelen. Diverse organisaties gebruiken de techniek voor verschillende doeleinden. Bijvoorbeeld voor marketing, klantenservice, sollicitaties, onderwijs, publieke veiligheid en de gezondheidszorg. De markt voor het monitoren van stress en het voorkomen van een burn-out groeit ook hard. Emotieherkenning kan ook worden ingezet zonder dat je het doorhebt. Denk aan de inzet van emotieherkenning voor advertenties en marketing, bijvoorbeeld om koopgedrag in te schatten en te beïnvloeden.

De ontwikkeling en inzet van emotieherkenning groeit, omdat organisaties en individuen denken dat het grote voordelen op kan leveren. Organisaties gebruiken herkende emoties om producten, diensten en gezondheid te verbeteren, of voor de veiligheid. Bovendien hebben wearables privégebruik makkelijker gemaakt, bijvoorbeeld voor het bijhouden van stress. Ook zien bedrijven het belang van emotieherkenning voor betere interactie met klanten. In hoofdstuk 4 worden verschillende voorbeelden van emotieherkenning besproken. Verschillende weten-schappelijke artikelen noemen de mogelijke voordelen van emotieherkenning. Zo beschrijven wetenschappers bijvoorbeeld hoe systemen de zorg kunnen verbeteren door emoties van patiënten inzichtelijker te maken. Dit zou het anticiperen op de behoeftes van de patiënt makkelijker kunnen maken en de zorglast kunnen verminderen. Ook wordt emotieherkenning ingezet om personen met autisme te ondersteunen in sociale interacties. Zowel onderzoek naar emotieherkenning, als de markt voor de systemen, is dus toegenomen.

Tegelijkertijd twijfelen experts aan de basis van emotieherkenningssystemen en waarschuwen voor risico's op inbreuk van grondrechten en publieke waarden. Tijdens inzet en de ontwikkeling ontstaan risico's op inbreuk van grondrechten en publieke waarden, zoals privacy, menselijke waardigheid en non-discriminatie. Maar er zijn ook twijfels over twee omstreden aannames die de basis zijn van de systemen: meetbaarheid en universaliteit.

Box 3.1

Wat bedoelen we met 'emotieherkenningssysteem'?

Dit hoofdstuk gaat over biometrische emotieherkenningssystemen. Dit zijn AI-systemen die als doel hebben emoties te herkennen aan de hand van biometrie. Denk hierbij bijvoorbeeld aan hartslag, stemgeluid, maar ook postuur, geur en DNA. In box 4.1 is te lezen wanneer het gaat over 'biometrische gegevens'. Het type systemen in dit hoofdstuk wordt in de wetenschap op verschillende manieren genoemd. Dit komt omdat verschillende disciplines kijken naar emoties. In technische literatuur vallen de systemen onder 'affective computing'. Dit hoofdstuk kijkt niet naar sentimentanalyse: dat zijn AI-systemen die emoties afleiden uit tekst. Dat soort systemen wordt in wetenschappelijke literatuur vaak wel gevat onder emotieherkenningssystemen.[90]

De term 'emotieherkenningssysteem' staat ook in de AI-verordening[91], maar deze themastudie kijkt breder dan de verordening. Het hoofdstuk gaat over alle systemen die op basis van biometrie emoties proberen te herkennen. Ook zet deze rapportage verschillende risico's voor inbreuk op grondrechten en publieke waarden uiteen. Zo wordt een overkoepelend beeld geschetst.

Het hoofdstuk behandelt ook stress. Stress wordt vaak omschreven als een emotionele toestand, in plaats van een emotie zoals blijdschap, verdriet, of schaamte. Toch wordt stress vaak gebruikt als een indicator van bepaalde emoties en dus is het relevant voor emotieherkenningsonderzoek.

Het gebruik van de term 'emotieherkenning' betekent niet dat de systemen ook echt emoties kunnen herkennen. Dit hoofdstuk bevat dan ook de serieuze twijfels over de werking van deze systemen.

3.2 Wankele basisprincipes van emotieherkenning

Een eerste principe is dat emotieherkenningssystemen voor iedereen dezelfde indeling van emoties gebruiken. Deze aanname dat voor iedereen dezelfde indeling geldt, is zeer omstreden. De basis van de aanname is dat er natuurlijke emoties zijn die iedereen hetzelfde ervaart en uitdrukt. Dit heet ook wel de 'universaliteitshypothese'. Vaak zijn emotieherkenningssystemen getraind op basis van westerse ideeën over emotie, die niet per definitie voor iedereen gelden.[92] In werkelijkheid zijn emoties vaak complex en contextafhankelijk.[93] Zo heeft cultuur een sterke invloed op hoe emoties worden ervaren, geuit en benoemd. Dat context belangrijk is, hoeft niet te betekenen dat mensen helemaal geen gemeenschappelijke emotionele delers hebben.[94]

Om emotieherkenningssystemen te maken, moet ten tweede worden aangenomen dat emoties te herkennen zijn aan de hand van algemene en goed meetbare signalen. Deze aanname is ook omstreden. Er wordt van uitgegaan dat er algemene (fysieke en fysiologische) signalen zijn die duiden op de aanwezigheid van bepaalde emoties. Denk hierbij aan een glimlach als uitdrukking van geluk, of een hoge hartslag als indicatie van angst. Ontwikkelaars gebruiken de relatie tussen een meetbaar signaal en de emotie. Het signaal is dus een 'proxy' van de emotie. Het idee dat een proxy voldoende zegt om één emotie aan toe te kennen, is controversieel.[95] Het kan een signaal zijn van meer dan één emotie.[96] Zo is een hoge hartslag niet altijd een indicatie van angst en is een hard stemgeluid niet altijd een uitdrukking van woede. Dit kunnen ook indicaties zijn van bijvoorbeeld meer positieve emoties. Dit maakt emotieherkenning niet voldoende betrouwbaar, specifiek en niet universeel toepasbaar.[97]

3.3 De training en werking van emotieherkenningssystemen

De twee omstreden aannames zijn terug te zien in de training en werking van de AI-systemen. Voor veel verschillende emotieherkenningssystemen komt de training en werking ongeveer overeen. In figuur 3.1 zijn de stappen van deze systemen schematisch weergegeven. Dit is een versimpeling van de werking van veel emotieherkenningssystemen.

Er bestaan verschillende soorten gegevens of "modaliteiten", die emotieherkenning kunnen gebruiken. Er zijn systemen die emoties herkennen op basis van stemanalyse. Zo kan snel en luid praten, duiden op boosheid of agressie. Andere systemen gebruiken gezichtsanalyse: emoties worden herkend in de stand van de mondhoeken, ogen en wenkbrauwen. Andere modaliteiten zijn hartritme, transpiratie, hersenactiviteit,

DNA en geur. Naast systemen die één modaliteit gebruiken, bestaan er ook multimodale systemen. Deze systemen combineren verschillende soorten biometrische gegevens. Bijvoorbeeld gebaren gecombineerd met hartslag.[98]

Het AI-systeem meet eerst data om emoties in te herkennen. Dit kan op verschillende manieren. Camera's en microfoons kunnen onder andere stemgeluid, gezichtsuitdrukking en houding vastleggen. Dit zijn bekende en veelgebruikte toepassingen van meetsystemen. Er bestaan ook wearables die hartslag en zelfs hersenactiviteit meten. Bijvoorbeeld smartwatches, ringen en koptelefoons. Deze worden steeds goedkoper en kleiner, en daarmee toegankelijker voor iedereen. Wearables worden op het lichaam gedragen. Er zijn ook meetsystemen die op afstand meten, zoals camera's met gezichtsherkenning.

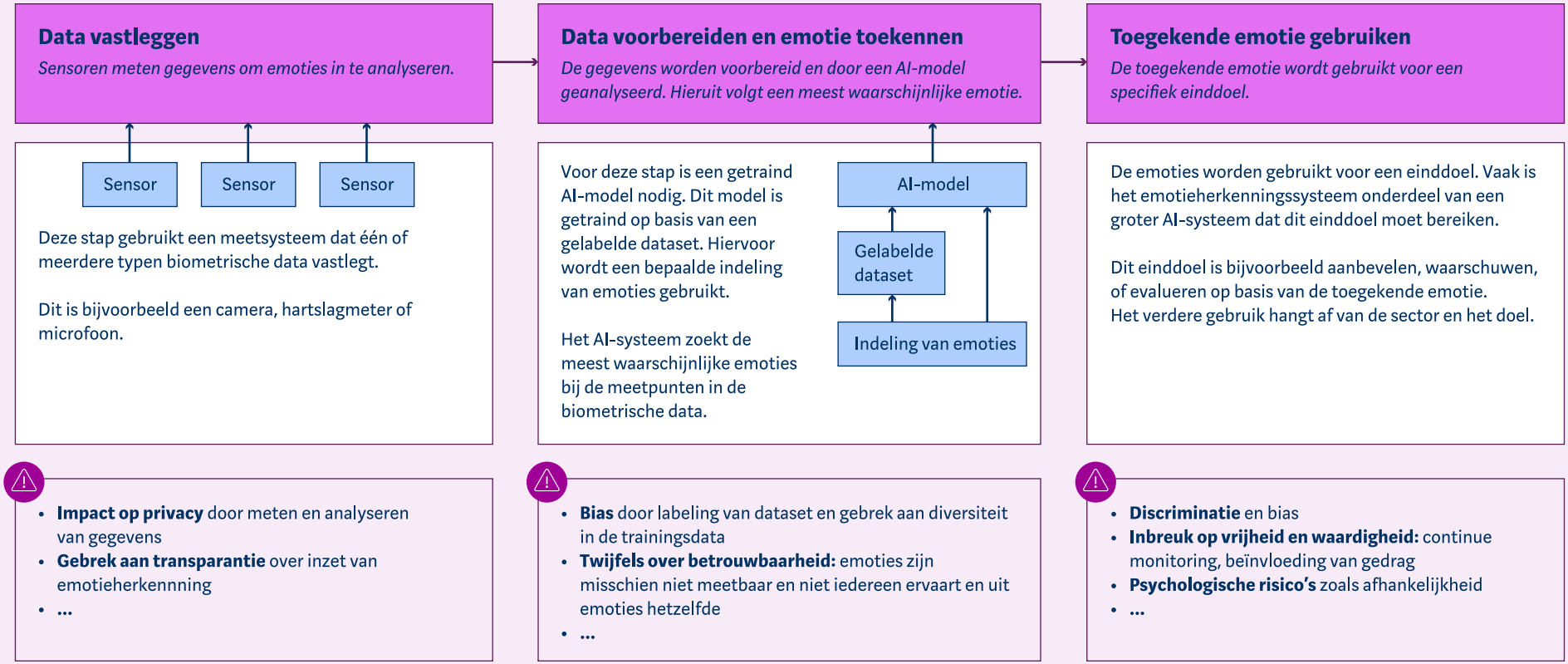
De gemeten data worden eerst voorbereid voor gebruik in het emotieherkenningssysteem. Tijdens de voorbereiding wordt een foto bijvoorbeeld bijgesneden, of er wordt ruis uit een stemopname verwijderd. Hierna selecteert het systeem vaak alleen bepaalde kenmerken, zoals gezichtsuitdrukkingen in een foto. Deze kenmerken worden op zo'n manier omgezet dat ze bruikbaar zijn voor het AI-systeem. Bijvoorbeeld door een code te gebruiken voor een opgetrokken wenkbrauw.[99] Alleen een abstractie van de data, de code, gaat het verdere AI-systeem in. De voorbereiding en omzetting van data wordt vaak ook door een AI-model gedaan. Een emotieherkenningssysteem bestaat dan eigenlijk uit meerdere AI-modellen.

Het AI-model analyseert de invoer en kent hier de meest waarschijnlijke emotie(s) aan toe op basis van een bepaalde indeling van emoties. Er bestaan meerdere indelingen. Veel ontwikkelaars van emotieherkenningssystemen gebruiken een versie van zes basisemoties: angst, woede, verdriet, walging, geluk en verrassing.[100] Soms zijn hier schaamte, schuldgevoel, trots, medeleven, opluchting, hoop en liefde aan toegevoegd. Deze indeling wordt veel gebruikt in AI-systemen, omdat het makkelijk te vertalen is in een algoritme.[101] Je kunt emoties hierbij ook verder indelen in een model met twee assen.[102] De eerste dimensie is of iets als positief of negatief wordt ervaren ('valence'). De tweede dimensie is de mate van opwinding ('arousal').[103] Een positieve emotie met neutrale opwinding is bijvoorbeeld blijdschap. En een negatieve emotie met weinig opwinding is verveling.

Het AI-model is hiervoor getraind met data waaraan 'labels' met emoties zijn toegevoegd. De manier waarop deze labels gemaakt zijn, verschilt per systeem. Als het model gezichtsuitdrukkingen gebruikt, bestaat de trainingsdata bijvoorbeeld uit foto's met een emotielabel. Labels kunnen op verschillende manieren worden gemaakt. Een testpersoon kan zelf aangeven welke emotie hij of zij ervaart, of iemand anders kan het emotielabel toeschrijven aan de data.[104] De data kunnen worden verzameld 'in het wild' of in het lab. In het lab worden soms emoties uitgelokt, bijvoorbeeld door bepaalde beelden te tonen. Ook kan iemand gevraagd zijn om een emotie te spelen. Dit hoeft dus niet de emotie te zijn die iemand echt ervaart. Dit alles bepaalt wat het systeem leert te voorspellen.

De toegekende emoties worden tenslotte verder gebruikt voor een specifiek einddoel. Vaak is een emotieherkenningssysteem onderdeel van een groter AI-systeem. Dan worden emoties verder gebruikt, bijvoorbeeld om aanbevelingen, evaluaties of waarschuwingen te geven. Hiervoor is soms nog meer data nodig over de context. Andere keren is het toekennen van een emotie het einddoel. Bijvoorbeeld in een app die emoties of emotionele toestanden als stress bijhoudt. Hoofdstuk 4 bespreekt verschillende toepassingen van emotieherkenningssystemen in de praktijk.

FIGUUR 3.1 | VERSIMPELDE WERKING VAN EMOTIEHERKENNINGSSYSTEMEN IN DRIE STAPPEN



Box 3.2

Emotieherkenning door narrow-AI versus general purpose AI

De weergegeven ontwikkeling en werking gaat over emotieherkenningssystemen die expliciet voor dit doel gemaakt zijn. Deze systemen worden ook wel narrow-AI genoemd. Ze zijn ontwikkeld voor specifieke taken in afgebakende domeinen. Ook gebruiken ze vooraf gekozen indelingen van emoties. Ze kunnen dan ook alleen emoties herkennen waarop ze getraind zijn.[105]

Daarnaast zijn general purpose AI-systemen in opkomst, die ook emotie-analyses kunnen uitvoeren, zonder expliciete training.[106] Deze systemen zijn dus niet specifiek gemaakt voor emotieherkenning. Het analyseren van emoties is een bijkomende vaardigheid van het model. Waar een narrow-AI-systeem bijvoorbeeld alleen de emoties kan herkennen waarop het getraind is, blijft een algemeen AI-systeem niet binnen één specifieke indeling. Ook kunnen ze toegepast worden op taken in verschillende domeinen.

Het is echter minder duidelijk hoe deze AI-systemen tot uitkomsten komen.[107] De analyses zijn niet op een specifieke manier getraind, maar komen voort uit de patroonherkenning op grote en diverse datasets. Op dit moment gaat het vooral om systemen die emoties kunnen analyseren in tekst (sentimentanalyse). In de afgelopen vijf jaar zijn er ook steeds meer systemen op de markt gekomen die foto's en geluid kunnen analyseren. Ook op basis van deze modaliteiten kunnen de systemen emotie-analyses doen.

Onderzoek naar emotieherkenning in tekst laat mogelijk inconsistente uitkomsten zien, doordat de grote datasets bijvoorbeeld veel verschillende soorten data bevatten.[108] Als deze systemen worden ingezet, kan dit zorgen voor misleidende resultaten en bias in de uitkomsten. Deze mogelijke inconsistentie is bovendien niet goed in te schatten, waardoor de resulterende bias moeilijk structureel te meten is. Het is voor algemene AI-systemen lastiger zulke bias te ontdekken of aan te pakken dan voor narrow-AI. De dataset is namelijk groter en niet specifiek gericht op emoties. Ook de training van het model is hier niet specifiek op gericht. Het is immers een bijkomende vaardigheid. Daarom zullen safeguards met name gericht zijn op prompts en benchmarks op de uitkomsten, die geregeld worden omzeild. Onderliggende bias in datasets wordt echter niet aangepakt.

General purpose AI-systemen beschrijven de uitkomsten vaak overtuigend, ook als het fout, incompleet of anderszins misleidend is. Het antwoord zal bijvoorbeeld een uitleg zijn over de herkende emotie, terwijl deze analyse verkeerd kan zijn. Door de uitleg wordt mogelijk meer waarde gegeven aan de uitkomst en sneller gedacht dat het model een goede voorspelling geeft.[109]

De afweging tussen enerzijds controle en anderzijds de gemiddelde kwaliteit zal bij emotieherkenning in toenemende mate van belang zijn. De verwachting is dat de gemiddelde kwaliteit van emotieherkenning in de toekomst hoger zal zijn bij systemen gebaseerd op algemene AI-systemen dan bij narrow-AI-systemen.[110] Bovengenoemde nadelen van inconsistentie, gebrek aan uitlegbaarheid, controleerbaarheid en overmatige overtuiging blijven echter staan.

Het volgende hoofdstuk gaat kort in op emotie-analyses door enkele grote taalmodellen. De AP deed hiervoor verkennende tests bij vier van zulke AI-systemen.

3.4 Risico's van emotieherkenning

Bij de inzet van emotieherkenningssystemen én gebruik van uitkomsten ontstaan risico's op inbreuk van publieke waarden en fundamentele rechten van burgers. Deze risico's kunnen ontstaan tijdens de verschillende stappen van de AI-systemen. Ook de ontwikkeling draagt bij aan het ontstaan van de risico's.

Zo kunnen de AI-systemen discriminerend werken door bias in de labels van trainingsdata. Vaak wordt het labelen van de trainingsdata voor AI-systemen met emotieherkenning als beoogd doel door mensen gedaan. Deze labels zijn dus de emoties die het systeem uiteindelijk moet herkennen. Vooroordelen die de labelaars hebben, worden mogelijk doorgegeven aan de trainingsdata en worden zo onderdeel van het AI-systeem. Zo speelt de bias ook door in de emoties die het systeem toekent. Het is dus niet alleen belangrijk wat het systeem leert, maar ook van wie.[111]

De systemen kunnen ook tot discriminatie leiden als de trainingsdata niet voldoende representatief zijn. Het model kan minder goed de juiste meetpunten selecteren en emoties toekennen aan een groep die weinig voorkomt in de dataset. Datasets zijn vaak onvoldoende divers qua cultuur, huidskleur, leeftijd en geslacht.[112] Studies hebben bijvoorbeeld laten zien dat biometrische systemen niet evengoed werken voor alle huidskleuren. Soms wijzen ze zelfs negatievere gevoelens aan zwarte mensen toe.[113] Ook ontbreekt vaak data over kinderen en ouderen, terwijl emoties en uitdrukkingen verschillen per leeftijdsgroep. Tenslotte kunnen neurodivergente personen, zoals mensen met autisme[114], en personen met gezondheidsproblemen afwijken van de trainingsdata.[115] Met de inzet van emotieherkenning kan iemand hierdoor bijvoorbeeld onterecht een slechte beoordeling krijgen op werk, of als ongeschikt worden aangemerkt tijdens een sollicitatie.

Emotieherkenningssystemen kunnen op basis van hun analyse suggesties of specifieke aanbevelingen doen. Dit kan de keuzevrijheid en autonomie van personen inperken. Aanbevelingen op basis van de toegekende emoties kunnen het gedrag van mensen bewust of onbewust beïnvloeden. Bijvoorbeeld als koopgedrag wordt beïnvloed op basis van vastgestelde emoties. Dit kan de keuzevrijheid aantasten. Zeker als dit op een manipulatieve of deceptieve manier gebeurt, is dit in strijd met de grondrechten.

Ook het constant monitoren van emoties door deze systemen kan inbreuk doen op menselijke waardigheid.[116] Als de inzet van de systemen bekend is, kan constante monitoring van emoties als vervelende en intieme surveillance worden ervaren. Dit kan bijvoorbeeld plaatsvinden op het werk, in het onderwijs en in (semi-) publieke ruimtes. Mensen kunnen hierdoor gestrester raken of de druk voelen om emoties niet te uiten of ze te verbergen.[117] Ook als de inzet van dergelijke systemen niet bekend is, kunnen ze afbreuk doen aan menselijke waardigheid, doordat mensen en hun emoties worden teruggebracht tot simplistische metingen van een systeem.

Ook privégebruik van emotieherkenningssystemen wordt steeds toegankelijker. Het is belangrijk dat gebruikers bewust omgaan met risico's. Wearables maken het makkelijker voor mensen om zelf emoties en emotionele toestanden zoals stress bij te houden. Apps geven ook tips om stress te verminderen en welzijn te verbeteren. De systemen kunnen hiermee ook een belangrijke rol krijgen in hoe mensen emoties ervaren. Hierdoor kan afhankelijkheid en overmatige zelfmonitoring ontstaan.[118] Ook kunnen mensen gaan twijfelen aan hun eigen oordelen. Het AI-systeem kan tussen de gebruiker en de eigen ervaring van emotie in komen te staan. Hierdoor begrijpen of beheersen mensen hun eigen emoties mogelijk minder zelfstandig. Mits de gebruiker zich bewust is van eventuele beperkingen, kan dergelijke technologie ook kansen en inzichten in emotioneel welzijn en gezondheid bieden.[119]

Het gebruik van biometrie door emotieherkenningssystemen zorgt voor risico's voor gegevensbescherming en privacy. De emotieherkenningssystemen gebruiken persoonlijke gegevens, zoals foto's, om emoties te analyseren. Ook de herkende emoties zijn zeer intieme informatie en dus privacygevoelig.[120] Het is dus belangrijk dat mensen weten welke gegevens emotieherkenningssystemen gebruiken én welke emoties worden geanalyseerd.

In sommige gevallen kunnen mensen door ongelijke machtsverhoudingen emotieherkenning lastig weigeren. Bijvoorbeeld bij emotieherkenning op werk of in het onderwijs. De machtsverhoudingen zijn hier ongelijk. Hierdoor is het niet altijd makkelijk voor werknemers en leerlingen om de inzet van emotieherkenning te weigeren. Mede hierdoor is de inzet van emotieherkenning op basis van biometrie in het onderwijs en op de werkvloer verboden onder de AI-verordening sinds 2 februari 2025.[121] Dit hangt samen met de twijfelachtige wetenschappelijk basis van dergelijke systemen. Ook in andere toepassingsgebieden kan weigeren van de inzet

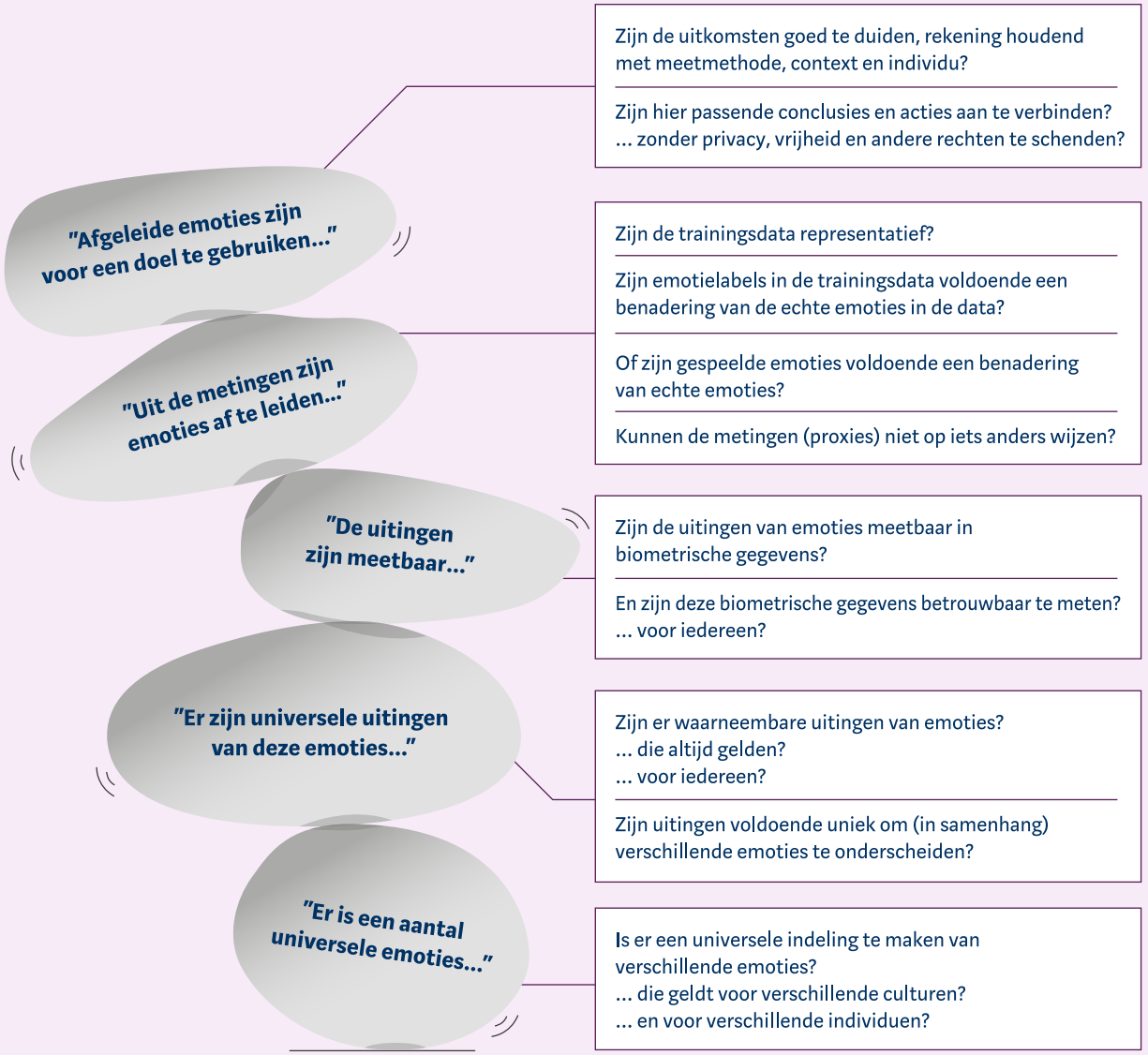
moeilijk zijn, zoals in de zorg, publieke ruimtes of bepaalde marketing contexten.

Het is dus ook belangrijk dat transparantie over de werking en inzet van dergelijke systemen goed op orde is, zodat mensen geïnformeerde beslissingen kunnen maken. Mensen kiezen vaak niet zelf om de systemen in te zetten. Naast de genoemde machtsverhoudingen, kan er dan ook sprake zijn van scheve kennisverhoudingen. Organisaties die emotieherkenningssystemen inzetten, weten meer dan de mensen van wie emoties worden herkend. Zij weten bijvoorbeeld niet of emoties op de achtergrond bijgehouden worden. Het herkennen van emoties is niet altijd het einddoel. Een persoon die een advies krijgt, weet dus niet meteen dat zijn of haar emoties gebruikt worden om tot dit advies te komen. Het is dus belangrijk om transparant te zijn over de inzet van emotieherkenning.

Emotieherkenningssystemen zijn in opkomst door veronderstelde voordelen. Tegelijkertijd roepen deze systemen fundamentele vragen op over de betrouwbaarheid, meetbaarheid en risico's voor grondrechten. De theorieën die de basis zijn van emotieherkenningssystemen vormen samen met de technische werking, de inzet en het gebruik een wankele toren. Deze is weergegeven in figuur 3.2. Deze aannames over emoties zijn omstreden, maar worden wel gebruikt in de systemen. In de ontwikkeling kan bias ontstaan. En de inzet van de systemen brengt verschillende risico's met zich mee. Om deze redenen dient men zeer voorzichtig met deze technologie om te gaan. Het volgende hoofdstuk gaat verder in op enkele toepassingsgebieden. Het bespreekt overige risico's, concrete voorbeelden in de praktijk en ook de mogelijke kansen van emotieherkenning.

FIGUUR 3.2 | BOUWSTENEN VOOR EMOTIEHERKENNINGSSYSTEMEN

De werking van AI voor emotieherkenning veronderstelt een wankele basis van aannames op vijf onderdelen.



4. Emotieherkenning in de praktijk

SNEL NAAR DIT ONDERDEEL

Praktijktest 1: Wearables →

Praktijktest 2: Taalmodellen →

Praktijkonderzoek 3: Klantenservice →

Hoe kom je in aanraking met emotieherkenning? De AP heeft drie toepassingsgebieden onder de loep genomen: **wearables, taalmodellen en de inzet in klantenservice**. Hieruit blijkt dat niet altijd duidelijk is hoe emoties en stress worden herkend. Ook is emotieherkenning niet altijd effectief. En hoewel AI met regelmaat wordt ingezet, is dit lang niet altijd duidelijk voor de persoon van wie een analyse wordt gemaakt. Daarnaast is technologische vooruitgang te zien. Apparaten worden kleiner en algoritmes en AI zijn in ontwikkeling. Organisaties en individuen zien kansen en een toename van het gebruik in de toekomst. In de praktijk is een risicobewuste weg vooruit erg belangrijk, gegeven risico's en de wankele theoretische basis waar emotieherkenning op is gebaseerd. Voor toezichthouders en beleidsmakers ligt er een opgave om te bedenken of de inzet van emotieherkenning in verschillende domeinen gewenst is. Het gaat om domeinen als de publieke ruimte, marketing, gezondheidszorg, dienstverlening en klantcontact.

Er worden kansen verondersteld in verschillende toepassingen van emotieherkenning, maar er zijn ook risico's. In figuur 4.1 staat een overzicht van enkele veelbesproken toepassingsgebieden. Hierin staan kansen, voorbeelden en risico's uit de literatuur. Dit maakt duidelijk hoe er naar de systemen gekeken wordt. Het overzicht biedt een kijk op verschillende toepassingen, maar is niet uitputtend.

De AP heeft gekeken naar drie toepassingsgebieden om inzicht te krijgen in emotieherkenning in de praktijk. De AP heeft enkele systemen getest in twee toepassingsgebieden: persoonlijke gezondheid en generatieve taalmodellen. Daarnaast is via een enquête inzicht verkregen in het gebruik van emotieherkenningssystemen binnen klantenserviceorganisaties in Nederland. Deze drie verdiepingen worden in dit hoofdstuk verder toegelicht.

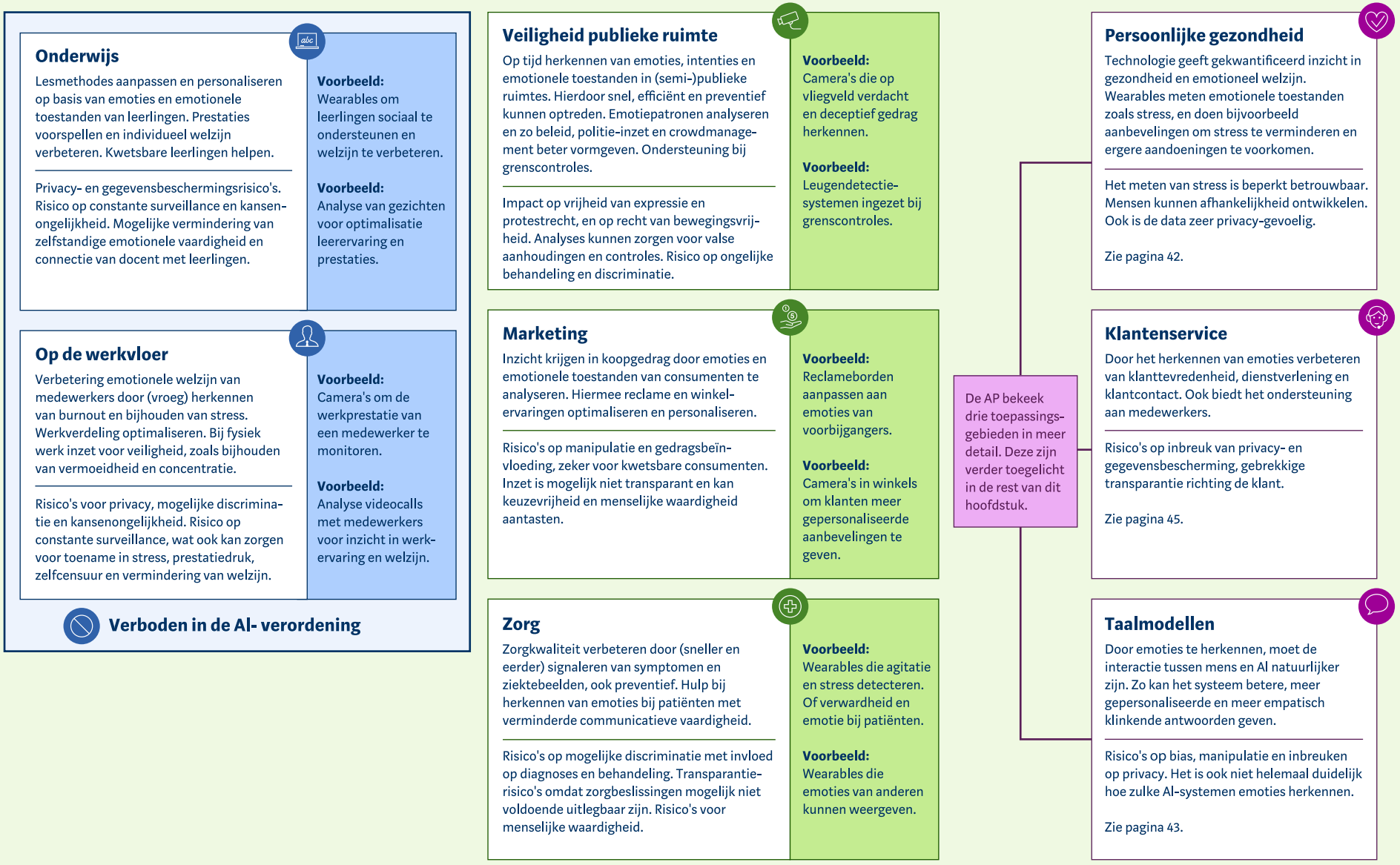
Een overkoepelende observatie is dat de inzet van emotieherkenning in de praktijk sector- en domeinspecifieke aandachtspunten meebrengt. Dat de inzet van emotieherkenningssystemen in verschillende domeinen risicovol is, vindt zijn weerslag in de AI-verordening. Voor het onderwijs en de werkvloer is zelfs expliciet vastgesteld dat de risico's zo groot zijn, dat het gebruik van emotieherkenning hier, op basis van die verordening, verboden is. In alle andere gevallen is emotieherkenning op basis van biometrie een hoog-risicosysteem onder de verordening.

Naast de beheersing van hoog-risicosystemen onder de AI-verordening is het belangrijk dat toezichthouders en beleidsmakers bedenken of inzet van emotieherkenning wenselijk is, ook waar dit niet verboden is. Het primaire doel van de beheersing van hoog-risicosystemen onder de AI-verordening is waarborgen dat deze systemen veilig zijn en voldoen aan productvereisten. De wenselijkheid van de inzet van deze systemen is echter een andere vraag. Het is uiteindelijk een politieke afweging om te bepalen in welke mate deze systemen wenselijk zijn en waarvoor deze systemen gebruikt mogen worden. Een analogie kan dit verduidelijken: productregulering helpt om te zorgen dat vuurwerk, aangeboden op de Nederlandse markt, veilig is. Maar de vraag waar en onder welke omstandigheden vuurwerk afgestoken mag worden, kent andere overwegingen. Een soortgelijk onderscheid tussen regulering en toezicht op het product enerzijds en het gebruik van dat product anderzijds, is ook van toepassing op emotieherkenning.

Ontwikkelaars en gebruikers moeten zich bewust zijn van de risico's die elke specifieke toepassing met zich meebrengt. Verantwoorde inzet van systemen vraagt om veiligheid en transparantie. Transparantie over deze analyse en de wijze waarop dat gebeurt richting de personen wiens emoties worden geanalyseerd. Bovendien is toestemming van de personen die geanalyseerd worden een belangrijke eerste stap voor organisaties om de systemen verantwoord in te zetten. In dit hoofdstuk worden drie praktijkvoorbeelden van emotieherkenning besproken. Maar ook voor andere toepassingen is werken aan een verantwoorde, bewuste en transparante inzet belangrijk.

FIGUUR 4.1 | OVERZICHT VAN TOEPASSINGSGEBIEDEN VOOR EMOTIEHERKENNING

Emotieherkenning in de praktijk: kansen en risico's in verschillende toepassingsgebieden.



Praktijktest 1: Wearables

Wearables zijn al lange tijd erg populair en bieden steeds meer opties. Dit zijn apparaten die je kan dragen, bijvoorbeeld om je pols. Met verschillende sensoren meten ze onder andere beweging, hartslag, temperatuur en slaap. Algoritmes kunnen patronen inzichtelijk maken en aanbevelingen doen.

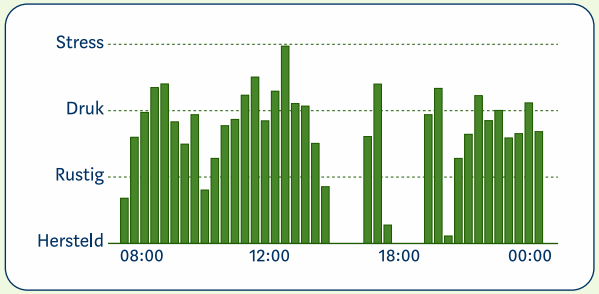
De AP keek naar drie wearables die stress meten om inzicht te krijgen in deze functies. De test is uitgevoerd met een horloge, een ring en een fitness-activiteits-tracker. Het gaat om producten die op de Nederlandse consumentenmarkt beschikbaar zijn. Alle wearables geven stress-scores weer en doen suggesties om stress te verlagen. In bijbehorende apps zijn hier oefeningen voor. Ook geven de apps inzicht in oorzaken van stress, zoals onvoldoende slaap en beweging.

Hoe wearables precies stress-scores berekenen is niet helemaal duidelijk en onderliggende data zijn niet altijd inzichtelijk. Hoe de verschillende indicatoren samen gebruikt worden om stress te detecteren, is niet duidelijk. De variabelen worden soms wel benoemd, maar de verbanden ertussen niet. Ook is het niet altijd mogelijk deze variabelen los te bekijken.

Alle apparaten gebruiken hartslagvariabiliteit als indicator voor stress, maar dit is niet altijd even betrouwbaar. Hartslagvariabiliteit is de tijd tussen hartslagen. Het hangt sterk samen met beweging. De apps geven tijdens beweging soms dan ook geen score weer. Of juist een erg hoge stress-score. Die hoge stress-score kan dus ook op beweging duiden.

De presentatie laat stress-scores er betrouwbaar en objectief uit zien. Stress wordt in een app weergegeven in duidelijke diagrammen met categorieën en heldere scores. Zo lijkt het betrouwbaar. Ook lijkt het objectief, terwijl de subjectieve kant van stress ook erg belangrijk is.

FIGUUR 4.2 | VERSIMPELDE WEERGAVE VAN DE WIJZE WAAROP EEN WEARABLE DE EMOTIEMONITORING ZICHTBAAR MAAKT IN APP



Toelichting: De app geeft aan: "Je had 3 uur last van stress, dus het was een stressvolle dag."

Inzichten, meldingen, adviezen en verschillende functies wekken veel aandacht naar de apps. Dit trekt gebruikers meermaals per dag naar de apps toe. Stress-scores kunnen soms meerdere keren per uur worden weergegeven. De apps geven meldingen met adviezen. Bijvoorbeeld om oefeningen uit de apps te doen.

Er komen ook steeds nieuwe functies beschikbaar voor hetzelfde apparaat, terwijl het niet nieuwe soorten data meet. Het gaat om nieuwe analyses op basis van dezelfde input. In algoritmes wordt input op nieuwe manieren gecombineerd. De hardware blijft hetzelfde, de software verandert. Ook zijn er functies zoals een chat en oefeningen, die losstaan van de metingen.

De verschillende functies zijn onderdeel van het verdienmodel achter de wearables. Alle geteste apparaten bieden abonnementen aan. Zonder abonnement zijn niet alle functies beschikbaar. Of de app geeft scores, maar geen uitgebreide analyses. Voor analyses en oefeningen voor betere gezondheid moet de gebruiker dus vaak betalen.

Als de gebruiker bewust omgaat met de wearables, kunnen ze op langere termijn nuttige inzichten geven. Over tijd leert het apparaat de gebruiker beter kennen. Hierdoor kan het meer accurate scores geven. Ook leert de gebruiker het apparaat beter kennen. Zo kan diegene beter begrijpen wat er gemeten wordt, maar ook wat de gebreken zijn en wat de scores voor hen betekenen.

Praktijktest 2: Taalmodellen

Emotieherkenning kan interacties met taalmodellen persoonlijker maken. Antwoorden sluiten zo beter aan bij emoties van de gebruiker. Een chatbot kan daarmee bijvoorbeeld beter inspelen op emoties.

Verschillende taalmodellen kregen het afgelopen jaar naast tekst-, ook spraak- en foto-opties. Daardoor kan je een live gesprek aangaan met het taalmodel. Ook kan je foto's van jezelf of je omgeving uploaden.

De AP deed een verkennende test met vier taal-modellen om te kijken naar emotieherkenning in beeld en geluid. In gesprekken werden dezelfde teksten op verschillende manieren opgezegd. Ook werden foto's ingevoerd met verschillende houdingen en gespeelde gezichtsuitdrukkingen. De systemen werd steeds gevraagd welke emoties het herkende. Resultaten op hoofdlijnen zijn weergegeven in tabel 4.1.

De verschillende modellen lijken geluid om te zetten in tekst en dit te analyseren. Eén model gaf aan ook veranderingen in toonhoogte, luidheid en snelheid te kunnen analyseren. Dit bleek echter niet uit de reacties van het model. Het leek vooral af te gaan op de inhoud van berichten. De drie overige taalmodellen gaven aan alleen tekst te analyseren, of gesproken woorden meteen om te zetten in tekst en dit te analyseren.

Aan foto's van gezichten kende drie van de vier modellen de 'gespeelde' emoties redelijk toe. In antwoorden benoemden ze vooral ogen, mond, wenkbrauwen en spanning in het gezicht. Op basis hiervan stelden ze verschillende mogelijke emoties voor. In figuur 4.3 staat hiervan een voorbeeld.

Ook konden de modellen houdingen analyseren, maar hier slechts beperkt emoties aan toekennen. Ze konden de houding benoemen en drie modellen koppelden hier ook een betekenis aan. Een houding met armen boven het hoofd is 'enthousiast' of 'triomfantelijk', maar kan ook 'een vraag om hulp' zijn. De uitleg hangt erg af van de herkende gezichts-uitdrukking in dezelfde foto.

De applicaties bouwen een overtuigend verhaal op basis van de input. Hierin tellen ook de omgeving, kleding en accessoires mee. Een model merkte bijvoorbeeld op dat foto's in een kantoor waren genomen. Mede hierdoor concludeerde het model dat de emoties mogelijk gespeeld waren. Een voorbeeld van zo'n verhaal staat in figuur 4.4.

Hoe emoties precies worden toegekend, blijft onduidelijk. In antwoorden verwijzen de modellen dus naar verschillende kenmerken die ze gebruiken. Maar hoe ze deze gebruiken, wordt niet helemaal duidelijk. De modellen benoemen bijvoorbeeld dat ze de kenmerken 'bestuderen', 'interpreteren' of 'patronen in visuele data' analyseren.

TABEL 4.1 | MODALITEITEN DIE AI-TAALMODELLEN GEBRUIKEN OM EMOTIES TE HERKENNEN*

Modaliteit	Taal-model 1	Taal-model 2	Taal-model 3	Taal-model 4
Stem (audio)	niet herkend/ onbekend	niet herkend	niet herkend	niet herkend
Gezicht (foto)	wel herkend	wel herkend	wel herkend	niet herkend
Houding (foto)**	wel herkend	wel herkend	wel herkend	niet herkend

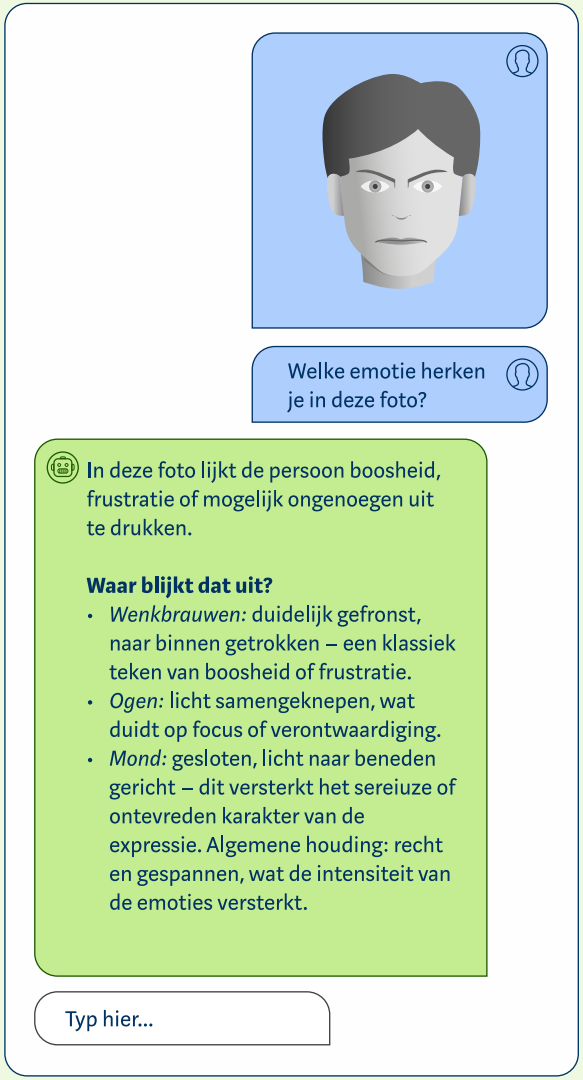
* Tijdens uitvoering praktijktest (april-mei 2025)
** Alle modellen konden in enige mate houding analyseren. Drie konden het in relatie tot emotie, maar gezicht leek alsnog leidend.

Wel verwijzen twee modellen naar omstreden emotietheorieën, die blijkbaar in trainingsdata aanwezig zijn. In hoofdstuk 3 werden deze theorieën al benoemd. Ze zijn veelbesproken en waarschijnlijk sterk aanwezig in de trainingsdata. Daarom gebruiken de modellen deze theorieën waarschijnlijk ook. Zie bijvoorbeeld figuur 4.4. In dat geval worden gebreken en bias van de theorieën doorgegeven aan het AI-model.

Risico's ontstaan als zulke modellen verder worden gebruikt. Organisaties kunnen met de taalmodellen nieuwe systemen bouwen. Deze systemen kunnen dan ook emoties herkennen, ondanks dat emotie-herkenning misschien niet direct het doel is van de nieuwe systemen.

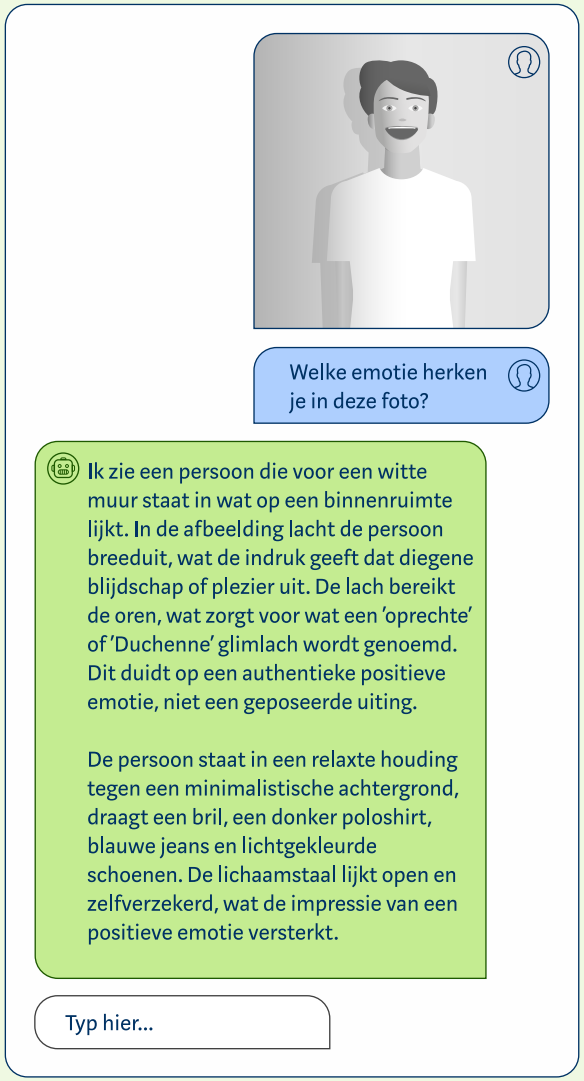
Personalisatie op basis van herkende emoties kan bij privégebruik tot risico's leiden. Een chatbot kan bijvoorbeeld manipulatieve effecten hebben of verslavend werken. Mensen kunnen afhankelijk worden, bijvoorbeeld als ze taalmodellen voor vriendschap en therapie gebruiken. De risico's hiervan zijn in de Rapportage AI- en Algoritmerisico's Nederland (RAN) van voorjaar 2025 toegelicht. Het gebruik van taal-modellen voor deze doelen maakt de risico's acuter voor een nog groter publiek.

FIGUUR 4.3 | VOORBEELD VAN EMOTIEHERKENNING DOOR TAALMODEL



Toelichting: Passage uit test met een taalmodel, uitgevoerd voorjaar 2025.

FIGUUR 4.4 | TAALMODEL VERWEEFT EMOTIETHEORIEËN IN REACTIE OP VRAAG NAAR HERKENDE EMOTIES



Toelichting: Passage uit test met een taalmodel, uitgevoerd voorjaar 2025.

Praktijkonderzoek 3: Klantenservice

De klantenservice speelt een belangrijke rol in de relatie tussen klant en bedrijf. Goed begrip van de emoties van klanten kan de klantrelatie verbeteren. Een klantenservicemedewerker kan zo beter inspelen op de emoties van een klant en daarmee een optimale klantervaring bieden. Dit maakt de klantenservice persoonlijker en vaak effectiever. Daarom heeft de AP in samenwerking met de Klantenservice Federatie een enquête onder Nederlandse organisaties met *inhouse* en *facilitaire* klantenservice en leveranciers uitgezet. Het doel was om inzicht te krijgen in het (toekomstig) gebruik van emotieherkenningssystemen voor de optimalisatie van klantcontact.

De AP kreeg via een enquête inzicht in het gebruik van emotieherkenningssystemen op basis van biometrie in klantenservices. Ongeveer dertig organisaties, zowel leveranciers als organisaties met *inhouse* of *facilitaire* klantenservice, gaven inzicht in hun kijk op het gebruik van emotieherkenning in klantcontact. De vragen gingen over het (potentiële) gebruik, de kansen en de risico's, en de verwachte, toekomstige inzet van emotieherkenningssystemen in klantcontact in Nederland.

De enquête onder organisaties met *inhouse* of *facilitaire* klantenservice toont aan dat 45% van de respondenten potentie ziet in emotieherkenning voor klantcontact. Een van de facilitaire klantcontactorganisaties heeft tussen de 50-249 klantcontactmedewerkers in dienst en past emotieherkenning al toe via spraakopnames (zie grafiek 4.1). Deze organisatie beoordeelt dit als enigszins effectief. Een kleinere organisatie die emotieherkenning ook al toepast, beoordeelt de effectiviteit neutraal. Organisaties die aangeven deze systemen in te zetten, geven ook aan te werken aan AI-geletterdheid. Medewerkers krijgen de nodige training of educatie over de inzet van zulke systemen.

Het aantal klantenserviceorganisaties dat aangeeft al emotieherkenning te gebruiken is dus klein. Daarbij bestaat ook de kans dat organisaties nog niet volledig beseffen dat hun systemen emotieherkenning bevatten, bijvoorbeeld omdat het niet als zodanig wordt bestempeld. Ook kan het zijn dat organisaties deze toepassingen niet vanzelfsprekend interpreteren als AI. Dit is relevant, omdat deze toepassingen straks wel onder specifieke AI-regelgeving vallen en ook nu aan AVG-vereisten moeten voldoen, vanwege de persoonsgegevens die verwerkt worden.

Emotieherkenning biedt volgens de respondenten verschillende kansen. Een groot deel van de respondenten is van mening dat emotieherkenning kan bijdragen aan klanttevredenheid en het verbeteren van de kwaliteit van de dienstverlening. Organisaties zien ook het voordeel in het ondersteunen en beschermen van medewerkers. Een enkeling zegt daarbij dat emotieherkenning ook ingezet kan worden als nieuwe bron voor klantdata. Het doel kan daarbij zijn om dienstverlening te personaliseren. Bovendien dient emotieherkenning in klantenservice als kwaliteitsmonitoring van gesprekken. Eén van de respondenten vraagt zich wel af in hoeverre technologie hier een zinvolle toevoeging is. 'Medewerkers zijn toch zelf prima in staat om emoties te herkennen?'

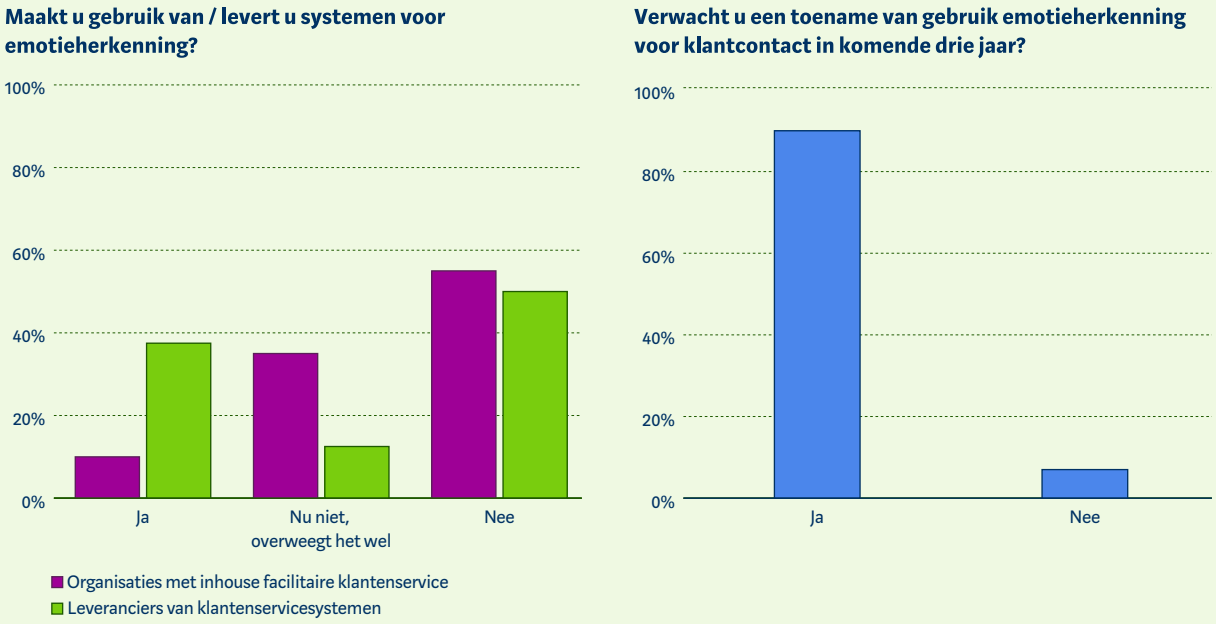
Het merendeel van de klantenserviceorganisaties is zich ook bewust van de risico's. Respondenten geven aan dat de inzet van emotieherkenning onbetrouwbaar kan zijn en kan zorgen voor bias en het gevoel van excessieve controle bij medewerkers. Meerdere organisaties zijn zich ook bewust van het risico op inbreuk van de privacy van klanten. Eén van de respondenten (leverancier) maakt duidelijk dat er geen risico's zijn als de toepassing van emotieherkenning een real-time ondersteuning is, en daarmee zinvol en onherleidbaar. Bij dergelijke situaties kunnen echter nog steeds risico's ontstaan. Bij de verwerking van persoonsgegevens of biometrische gegevens moet te allen tijde voldaan worden aan de AVG, de AIV en andere relevante wet- en regelgeving.

Enkele respondenten benadrukken dat risico's groter zijn bij de toepassing op individueel niveau dan bij toepassing op geaggregeerd niveau. Een respondent geeft aan dat emotieherkenning kan worden ingezet om anoniem en op geaggregeerd niveau conclusies te trekken. Hiermee menen zij de kwaliteit van de dienstverlening te kunnen verbeteren. Emotieherkenning kan ook op individueel niveau worden ingezet om de dienstverlening te personaliseren. Volgens respondenten brengt dit echter te veel risico's met zich mee. Ook op

geaggregeerd niveau is het raadzaam om de mogelijke risico's en de wettelijke eisen in acht te nemen.

De helft van de ondervraagde leveranciers van klantenservicesystemen geeft aan emotieherkenning aan te bieden of dit te overwegen. Zie grafiek 4.1. De voornaamste drijfveer voor de ondervraagde leveranciers is het verbeteren van de klanttevredenheid en de dienstverlening. Daarbij wordt gesteld dat inzicht in de emoties van klanten via emotieherkennings-systemen ook een bijdrage kan leveren in het sneller afhandelen van klachten. De leveranciers geven aan dat de emotie zowel tijdens als na het klantcontact wordt herkend.

GRAFIEK 4.1 | ONTWIKKELINGEN IN GEBRUIK EMOTIEHERKENNING BINNEN KLANTCONTACT



Bron: Eigen onderzoek onder organisaties met inhouse of facilitaire klantenservice en leveranciers van klantenservicesystemen. Enquête opgezet in samenwerking met Klantenservice Federatie (n=29) en uitgevoerd gedurende april-mei 2025.

Leveranciers en de klantcontactorganisaties geven aan dat emoties voornamelijk worden gemeten via spraakopnames. Emoties worden daarbij afgeleid uit verschillende gegevens: stemgeluid, intonatie of volume en ook door bijvoorbeeld pauzes of stiltes. Zie ook figuur 4.5.

Het gebruik van emotieherkenning tijdens klant-contact wordt door respondenten niet direct kenbaar gemaakt. Respondenten geven een algemene melding of geven alleen uitleg als een klant ernaar vraagt. Transparantie over het gebruik van algoritmes en AI en het gebruik van biometrische gegevens is belangrijk. Zo hebben klanten een keuze over het delen van hun gegevens en inzicht in het gebruik. Daarmee kan het risico op een eventueel oneerlijke behandeling, door de uitkomst van een emotieherkenningssysteem, het best gemitigeerd worden. Toestemming voor het gebruik van biometrische gegevens voor emotie-herkenning en een duidelijke uitleg over het doel ervan zal in veel gevallen een eerste stap zijn. Daarbij moet de uitleg concreter zijn dan bijvoorbeeld enkel een algemeen begrip als "kwaliteits- en trainingsdoeleinden".

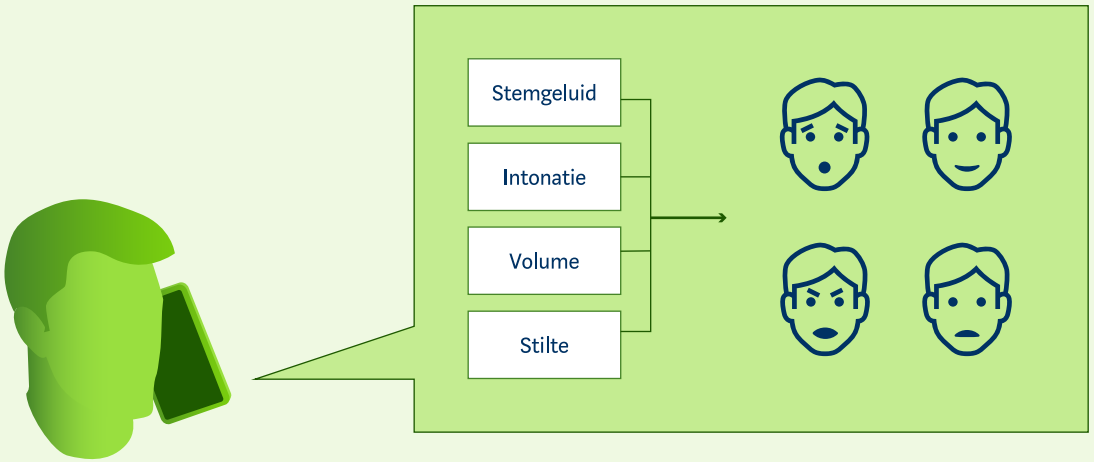
Een deel van de organisaties geeft aan te waken voor juridische compliance risico's bij verkeerde inzet. Bovendien acht meer dan 80% van de respondenten het opstellen van duidelijke richtlijnen over emotie-herkenningssystemen belangrijk tot zeer belangrijk. Dit geeft aan dat er inzicht is in de risico's, maar ook dat het belangrijk is om kennis te delen en training te bieden voor de inzet van zulke systemen.

Tegelijkertijd verwacht 90% van de respondenten een toename in het gebruik van emotieherkennings-systemen door Nederlandse klantcontact-organisaties. Een deel daarvan acht de kans ook groot dat hierbij gebruik wordt gemaakt van verschillende modaliteiten. Een enkeling verwacht terughoudendheid of afbouw.

Het is essentieel dat organisaties zich kritisch verhouden tegenover emotieherkenningssystemen. Hierbij moeten zij de wankele basisprincipes, technische werking en bijbehorende risico's van zulke systemen in acht nemen. Wie deze systemen ontwikkelt en inzet, doet er goed aan om duidelijk te stellen wat de meer-waarde van de technologie is om het beoogde doel te bereiken.

Bij de inzet van emotieherkenningssystemen is het cruciaal om transparant zijn over het gebruik en de werking van deze systemen, en de verzameling van bepaalde gegevens. Bovendien moeten medewerkers die gebruikmaken van de systemen de nodige kennis en kunde hebben. Het is daarom van groot belang dat organisaties AI-geletterdheid op orde hebben.[122] Zo moeten medewerkers onder andere begrip hebben van de risico's. Bovendien moeten organisaties die emotieherkenning inzetten via AI-systemen voldoen aan de hoog-risicovereisten in de AI-verordening.

FIGUUR 4.5 | WERKWIJZE VAN EMOTIEHERKENNING BINNEN KLANTCONTACT



Toelichting: Emoties worden geanalyseerd op basis van verschillende gegevens.

Box 4.1

Regelgeving biometrische gegegevens: inzoomen op AI-systemen voor emotieherkenning op basis van biometrie

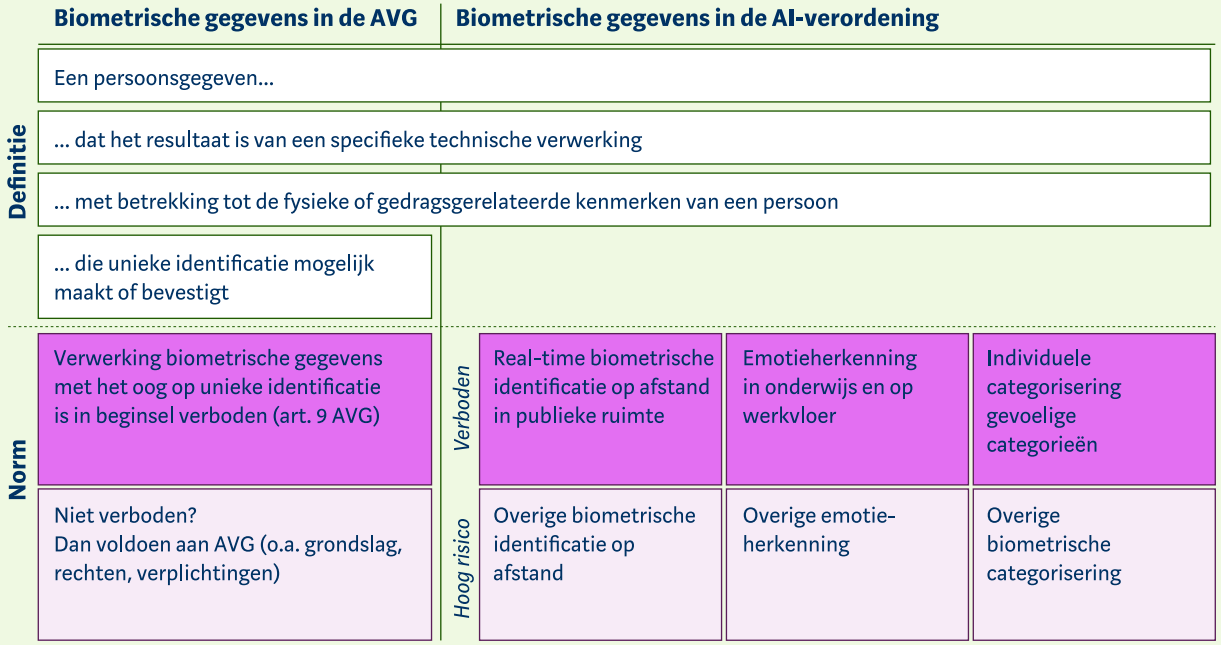
De inzet van emotieherkenningstechnologie op basis van biometrie raakt aan zowel de AI-verordening als de AVG.[123] Deze box geeft inzicht in het begrip 'biometrische gegevens' in beide verordeningen.[124] Het uitgangspunt is dat het begrip in de AI-verordening uitgelegd moet worden in het licht van de AVG. Maar uit de reacties op de oproep tot input van de AP over het verbod op emotieherkenning in de AI-verordening blijkt dat hier onder belanghebbenden verwarring over bestaat.[125] Dit vraagt om verdere verduidelijking, ook op Europees niveau.[126]

De AI-verordening benoemt dat biometrische gegevens in het licht van het gelijknamige begrip in de AVG moet worden begrepen.[127] Beide verordeningen hebben vergelijkbare doelen: de bescherming van grondrechten en de bevordering van de interne markt. Ze vullen elkaar ook aan waar het gaat om de regulering van emotieherkenningssystemen. In beide definities van biometrische gegevens gaat het namelijk in de basis steeds om (1) persoonsgegevens, (2) die het resultaat zijn van een *specifieke technische verwerking* (3) met betrekking tot de fysieke, fysiologische of gedrags-gerelateerde kenmerken van mensen.[128]

De AVG bevat daarnaast ook de vereiste dat op grond van die persoonsgegevens eenduidige identificatie van die natuurlijke persoon mogelijk is of wordt bevestigd. Er moet dan dus sprake zijn van een gegeven dat het resultaat is van een specifieke technische verwerking, met betrekking tot iemands fysieke, fysiologische of gedragsgerelateerde kenmerken, waardoor je deze persoon kunt identificeren of diens identiteit kunt bevestigen.

Biometrische gegevens met het oog op de eenduidige identificatie van een persoon zijn bijzondere persoons-gegevens onder de AVG. De verwerking ervan brengt aanzienlijke risico's met zich mee voor de schending van grondrechten en voor publieke waarden. Daarom is het verwerken van deze gegevens extra beschermd onder de AVG: je mag deze alleen verwerken als aan specifieke voorwaarden wordt voldaan.[129] Biometrische data zijn bijvoorbeeld referentiedata in databases voor gezichts-herkenning, of de analyse van meetgegevens over iemands 'loopje', als die het mogelijk maken iemand te

FIGUUR 4.6 | BIOMETRISCHE GEGEVENS IN DE AVG EN DE AI-VERORDENING



identificeren. Een eenvoudige foto van iemand, of een video waarop iemand wandelt, is op zichzelf geen biometrisch gegeven. Het gaat daarbij namelijk niet om de technische verwerking ervan met als doel iemand eenduidig te identificeren of iemands identiteit te bevestigen.

De AI-verordening biedt onder andere bescherming voor AI-systemen die emoties herkennen en mensen categoriseren op basis van biometrische gegevens. De regels in de AI-verordening zijn risicogebaseerd en afhankelijk van de beoogde toepassing. Zo zijn AI-systemen voor emotieherkenning op basis van biometrische gegevens verboden in het onderwijs en op de werkvloer.[130] Ook verboden zijn bijvoorbeeld AI-systemen die mensen individueel in gevoelige categorieën indelen op basis van biometrische gegevens, zoals politieke opvattingen of seksuele gerichtheid.[131] Zulke AI-systemen mogen niet op de markt worden gebracht of gebruikt. Tegelijkertijd worden sommige andere AI-systemen op basis van biometrische gegevens als hoog-risico beschouwd. Die moeten voldoen aan specifieke producteisen. Denk hierbij aan emotieherkenning in alle andere situaties dan buiten de werkvloer en het onderwijs. Figuur 4.6 laat dit zien.

AI-systemen voor emotieherkenning vallen onder de reikwijdte van de AI-verordening als deze gebruik maken van biometrische gegevens met als doel het afleiden van emoties. Denk bijvoorbeeld aan gedrags- en beweegbiometrie of analyse van elektrocardiogrammen (ECG). Gebruik van een ECG door een AI-systeem is op zich geen gebruik van biometrische gegevens. Maar als ECG's, eventueel in combinatie met andere datapunten, in een AI-systeem worden geanalyseerd om tot vaststellingen te komen over iemands emotie, dan is sprake van emotieherkenning in de zin van de AI-verordening.[132]

Bijlage
Aan de slag met algoritmeregistratie

Over dit document

Als coördinerend toezichthouder op algoritmes en AI helpt de AP organisaties met het beheersen van AI- en algoritmerisico's. Deze bijlage geeft organisaties een aantal handvatten om aan de slag te gaan met algoritmeregistratie. Daarbij geldt dat onderzoek, beleidsvorming en standaarden op het gebied van algoritmes en AI nog volop in ontwikkeling zijn.

Organisaties in zowel de publieke als private sector zijn steeds meer afhankelijk van algoritmes voor allerlei processen en toepassingen. Algoritmes opnemen in een register is een goed begin om de risico's te beheersen die met algoritme-inzet gepaard gaan. De AP ziet dat veel organisaties het als een uitdaging zien om hiermee te beginnen. In deze bijlage geeft de AP acht concrete handvatten om zelf met algoritmeregistratie aan de slag te gaan. Ook geeft de AP een visie van het belang van algoritmeregisters voor het werk van toezichthouders (zie box 2).

1. Waarom is algoritmeregistratie belangrijk?

Een algoritmeregister bevat informatie over de inzet van algoritmes. Het is een soort logboek of database waarin informatie staat over het doel, de trainingsdata en wie (intern) verantwoordelijk is voor het algoritme. Een register kan voor één organisatie zijn of voor meerdere organisaties tegelijk, bijvoorbeeld een gezamenlijk register voor de onderwijssector. Ook kan een register verschillende verschijningsvormen hebben: het kan een Excel-bestand zijn maar ook een publiek toegankelijke website, zoals bij het Algoritmeregister van de Nederlandse overheid (zie box 1).

Op hoofdlijnen heeft algoritmeregistratie twee overkoepelende doelen. Het eerste doel is het bevorderen van interne controle, want het register maakt het algoritmegebruik inzichtelijk.

Het tweede doel is het bevorderen van externe controle door transparantie te bieden. Deze doelen zijn nevenschikkend (zie figuur 1).

Doel 1: Het bevorderen van interne controle op algoritmes (*governance*). Dit omvat onder andere de beheersing van de ontwikkeling, de implementatie, de verantwoordelijkheden, de risico's en de naleving van wet- en regelgeving rond het algoritmegebruik. Een

algoritmeregister is hiervan een belangrijk onderdeel, omdat deze een overzicht biedt van algoritmes met de daarbij behorende ontwikkeling, inzet en controle. Daarnaast blijkt uit onderzoek dat het opzetten van een register medewerkers in een organisatie in gang zet om actief na te denken over algoritmegebruik, met een 'disciplinerend effect' als gevolg.[133]

Doel 2: Het bevorderen van externe controle op algoritmes (*transparantie*). Gebrek aan transparantie kenmerkt veel problemen die bij de inzet van algoritmes een rol spelen, omdat door het gebrek hieraan de uitkomsten en naleving met wet- en regelgeving lastig te controleren zijn. Het opnemen van algoritmes in een extern toegankelijk register draagt bij aan transparantie omdat eindgebruikers, journalisten, toezichthouders en andere personen op wie een algoritme betrekking heeft, inzicht kunnen krijgen in de werking van een algoritme en deze kunnen controleren. Bovendien bevordert een openbaar register kennisuitwisseling tussen organisaties over het gebruik en de beheersing van algoritmerisico's. Zie voor het toezichtsperspectief ook box 2.

FIGUUR 1 | DOELSTELLINGEN ALGORITMEREGISTRATIE

Het bevorderen van interne controle op algoritmes (governance)

Het bevorderen van externe contole op algoritmes (transparantie)

Er komt een Europese database waarin nieuwe of gewijzigde hoog-risico AI-systemen vanaf augustus 2026 geregistreerd moeten worden zodra deze op de markt zijn gekomen en een CE-keurmerk hebben. Dit geldt voor aanbieders van AI-systemen voordat ze op de markt uitkomen en voor gebruikers van AI-systemen in de publieke sector (artikel 49 AI-verordening). Deze database kan overlap hebben met een algoritmeregister, maar is geen vervanging voor het inrichten van een eigen algoritmeregister. De database is voor hoog-risico AI-systemen volgens de verordening met een CE-keurmerk, terwijl een algoritmeregister juist kan gaan over het gebruik van alle soorten algoritmes. Een algoritmeregister kan dus ook gaan over algoritmische processen die niet voldoen aan de definitie van een AI-systeem volgens de AI-verordening. Deze algoritmes kunnen evengoed hoog risico met zich meebrengen. Daarnaast is een algoritmeregister in beginsel vormvrij. Registratie in het Europese register is dus ook niet hetzelfde als registratie in het algoritmeregister van de Nederlandse overheid.

Algoritmes kunnen ook geregistreerd worden in een verwerkingsregister voor persoonsgegevens. Organisaties houden ook een verwerkingsregister bij voor bepaalde risicovolle verwerkingen, zoals bij bijzondere persoonsgegevens (artikel 30 AVG, artikel 24 RGR). Een verwerkingsregister is bedoeld om verwerkingen van persoonsgegevens te kunnen verantwoorden. Algoritmes die dit doen kunnen hierbij dus meegenomen worden.

Of algoritmeregistratie daadwerkelijk bijdraagt aan het tegengaan van risico's hangt af van de implementatie. Registratie is pas waardevol als er ook maatregelen worden genomen voor de controle op algoritmes, zoals audit- of governance-maatregelen. Ook is de kwaliteit van de registratie zelf van belang. Zo kunnen registraties die gebrekkige en onduidelijke informatie bevatten weinig bijdragen aan het voorkomen van algoritmerisico's, incidenten en, in het ergste geval, bijdragen aan schijnveiligheid.[134]

Box 1

Het Algoritmeregister van de Nederlandse overheid

Het ministerie van Binnenlandse Zaken en Koninkrijkrelaties lanceerde in 2022 een centraal algoritmeregister om meer inzicht te geven in de inzet van impactvolle algoritmes in het publieke domein. Voorbeelden van impactvolle algoritmes die in het register horen zijn 'algoritmes die rechtsgevolgen hebben voor mensen, of die zorgen dat de overheid mensen classificeert'. Overheidsorganisaties kunnen informatie in het register registreren over wat het algoritme doet (algemene informatie, verantwoord gebruik en werking), onder welke categorie het algoritme valt (hoog-risico, impactvol of overig), of en welke impacttoetsen zijn uitgevoerd en wat de status van ontwikkeling is. De website van het Algoritme-register bevat een dashboard dat inzicht geeft in geregistreerde algoritmes. Er zijn inmiddels rond de 1000 algoritmes geregistreerd in het register. Het Algoritmeregister staat op: www.algoritmeregister.nl.

De Nederlandse overheid werkt aan een wettelijk kader voor het Algoritmeregister. Het wettelijk kader moet zorgen voor duidelijkheid over het verplicht registreren van algoritmes. Maar ook zonder deze verplichting is het Algoritmeregister nu al een waardevolle en praktische tool voor het bieden van transparantie en het bevorderen van controle op overheidsalgoritmes. Nog lang niet alle organisaties benutten het Algoritmeregister. De AP moedigt overheidsorganisaties aan om hier zoveel mogelijk gebruik van te maken.

Het Algoritmeregister kan worden ingebed in bestaande werkprocessen, bijvoorbeeld bij het contact met burgers. Ter illustratie: burgers kunnen een brief krijgen met daarin een besluit dat door middel van, of ondersteund door, algoritmes of AI tot stand is gekomen. Als de ontvanger van deze brief meer wil weten, dan kan in de brief doorverwezen worden naar het Algoritmeregister voor informatie over de werking van het algoritme dat betrokken was bij het besluit. Daarmee wordt het Algoritmeregister op een nuttige wijze in werkprocessen gebracht.

Een betrouwbare overheid biedt nuttige, relevante informatie op een toegankelijke manier. Deze informatie kan op verschillende manieren bij burgers terechtkomen. Bijvoorbeeld doordat burgers verplicht betrokken worden bij het ontwikkelen van een risico-volle verwerking van persoonsgegevens. Doordat bij beslissingen duidelijk gecommuniceerd wordt dat er een algoritme of AI-systeem gebruikt wordt. Of doordat er verwezen wordt naar het Algoritmeregister waarin informatie staat over het algoritme en de toepassing daarvan. Het positieve gebruik van tools en informatie kan organisaties helpen om verantwoord om te gaan met algoritmes en AI. Maar het stelt vooral burgers in staat om kennis te nemen van de inzet van algoritmes en AI, daarop door te vragen of de uitkomsten te betwisten. Dit draagt bij aan meer vertrouwen in de technologie en het gebruik daarvan in de publieke sector.

2. Acht handvatten voor algoritmeregistratie

Algoritmeregistratie roept nog vaak vragen op: wat registreer je, voor wie, en hoe gedetailleerd? Omorganisaties en sectoren te helpen bij het opzetten eninvullen van algoritmeregisters, geeft de AP acht praktische handvatten. Deze helpen bij het maken van keuzes over doel, inhoud en aanpak van registratie.

Handvat 1: Bepaal van tevoren het doel voor algoritmeregistratie. Het doel van algoritmeregistratie bepaalt dedoelgroep en welke inhoud en vorm het register zou moeten krijgen. Is een register met name bedoeld voor medewerkers om intern de governance van organisatie-algoritmes te verbeteren? Dan kan een Excel-bestand met organisatiejargon voldoende zijn. Is het doel om als financiële dienstverlener kredietbeoordelingsalgoritmes inzichtelijk en navolgbaar te maken voor klanten? Dan is een laagdrempelige en niet-technische uitleg in een register op een plek waar klanten deze kunnen raadplegen noodzakelijk.

FIGUUR 2 | ACHT HANDVATTEN VOOR ALGORITMEREGISTRATIE

Toelichting: Het opzetten van algoritmeregistratie vraagt om een structurele aanpak. Dit waarborgt dat een algoritmeregister toegankelijk en volledig is, op basis van een helder doel en scope, met de juiste transparantie en waarbij een goede verantwoordelijkheidsverdeling zorgt voor actualisatie en kwaliteitsborging. De acht handvatten maken het voor sectoren mogelijk dit op te pakken.

Handvat 2: Omschrijf nauwkeurig welke algoritmes binnen het doel van registratie vallen. Veel organisaties gebruiken talloze computerprogramma's die algoritmes gebruiken. Welke algoritmes in het register horen is dan ook een belangrijke vraag. Zo zijn er veel operationele en alledaagse algoritmes die niet per se in een register hoeven. Denk aan een applicatie die automatisch een afspraak in de agenda's zet (*auto-scheduling*), een spellingschecker of algoritmes die verkoopcijfers rangschikken. Het gekozen doel voor algoritmeregistratie is leidend voor welke algoritmes in het register moeten komen. Is het doel transparantie bieden aan ouders over algoritmegebruik in het onderwijs? Dan moeten in eerste plaats vooral algoritmes geregistreerd worden die voor hen van belang zijn. Bijvoorbeeld algoritmes die de schoolprestaties van hun kinderen beïnvloeden, zoals leerlingvolgsystemen. Registratie van algoritmes die niet voor deze doelgroep relevant zijn, draagt in zo'n situatie niet bij aan overzicht en transparantie, maar aan meer ruis. Andere overwegingen voor registratie kunnen zijn de complexiteit, impact of context van het algoritme. Hebben algoritmes bepaalde inherente risico's? Dan is het altijd nuttig deze in het register op te nemen. Zoals algoritmes die bepaalde risicoscores aan mensen toekennen.

Handvat 3: Start met de beschikbare informatie en ontwikkel daarna door. Met het (in)vullen van een register kunnen vragen opkomen over de werking van een algoritme, welke processen erbij betrokken zijn en hoe deze informatie op een zinvolle manier voor de doelgroep kan worden opgeschreven. Antwoorden formuleren op dit soort vragen is essentieel voor de controle op algoritmes, maar kan ook als uitdagend worden gezien en leiden tot uitstel. Stel in het begin daarom niet te hoge standaarden en start met de beschikbare informatie.

Is bijvoorbeeld alleen een korte beschrijving over de werking van het algoritme beschikbaar, dan is het een goed uitgangspunt om deze informatie alvast te registreren, in plaats van te wachten tot de informatie compleet is. Zelfs met minder verfijnde of incomplete informatie kan begonnen worden met registratie. Dit draagt bij aan het dialoog over een algoritme, dat ook kan bijdragen aan betere informatie. Tegelijkertijd moeten organisaties wel op termijn verfijning en aanvulling van informatie nastreven, want onvolledige informatie kan ook leiden tot onduidelijkheid. Een balans moet gevonden worden tussen de snelheid van registratie en hoe volledig of gedetailleerd deze wordt verricht.

Voorbeeld 1

Algorithmic Transparency Standard

Zeven Europese gemeenten die voorlopen met algoritmeregistratie, waaronder Amsterdam, Eindhoven en Rotterdam, hebben samen de *Algorithmic Transparency Standard* ontwikkeld. Deze standaard bevat een template met categorieën om met algoritmeregistratie aan de slag te gaan. De standaard is gericht op gemeenten, maar bevat ook nuttige informatie voor andere organisaties. De standaard staat samen met een guidance-document op www.algorithmregister.org.

Handvat 4: Ga verder dan de basiskenmerken en neem ook de context mee bij de registratie. Basiskenmerken voor de registratie van een algoritme zijn een beschrijving van de werking (beslisregels), wat het algoritme doet, wie er verantwoordelijk voor is en welke data en techniek er gebruikt worden. Maar net zo belangrijk als de basis-elementen zijn de achterliggende bestuurlijke, economische of beleidsafwegingen om het algoritme te gaan gebruiken. Dit is bijvoorbeeld nuttig om in de toekomst beter te begrijpen waarom het algoritme bestaat of, als blijkt dat het algoritme niet meer effectief is voor het behalen van de initiële doelstellingen. Zo kan het zijn dat een algoritme bepaalde wetgeving uitvoert of gebaseerd is op (wetenschappelijke) standaarden die niet meer actueel zijn. Met hulp van informatie uit het algoritme-register kan de beslissing om een algoritme aan te passen of uit te faseren beter onderbouwd worden.

Voorbeeld 2

Subsidieregeling met inkomensgrens

De fictieve gemeente Koornwoude heeft in 2018 een subsidieregeling opgezet om woningen meer hittebestendig te maken voor lagere inkomens. Koornwoude gebruikt een algoritme om aanvragen te filteren: inkomens hoger dan 32.000 euro per jaar worden automatisch afgewezen, ook al gaat het om een klein verschil. De gemeente vond het wenselijk om een harde inkomsensgrens te hanteren bij het ontstaan van de regeling vanwege beperkte budgetten en hulp aan lage inkomens.

In de jaren daarna verandert Koornwoude van beleid en vindt de gemeente het een prioriteit om meer maatwerk te bieden. In 2025 ontstaat er daarom veel verontwaardiging in de gemeenteraad, als blijkt dat veel subsidieaanvragen automatisch en zonder motivering geweigerd worden voor mensen met niet-hittebestendige huizen die prioriteit zouden moeten krijgen. Onderzoek in het algoritmeregister laat zien dat het criterium van de harde inkomensdrempel de oorzaak is, en dat dit criterium is gebaseerd op oude beleidsopvattingen.

Handvat 5: Maak de risico's en de daarbij genomen beheersmaatregelen inzichtelijk. De inzet van algoritmes kunnen risico's met zich mee brengen op gebieden als privacy, discriminatie en bias, misleiding, technologische soevereiniteit of cybersecurity. Besteed daarom extra aandacht in een register aan het in kaart brengen van risico's. Denk aan een beschrijving van de risico's met een risicoclassificatiesysteem, waarmee bijvoorbeeld algoritmes de classificaties laag, midden of hoog risico krijgen. Factoren die hierbij een rol spelen zijn bijvoorbeeld de complexiteit, autonomie, of er sprake is van menselijke tussenkomst en de impact van het algoritme. Ook is het gebruik van een algoritme in een hoog-risico gebied van de AI-verordening (annex 3) een goede indicatie of er sprake is van hoog risico, omdat algoritme-inzet in deze gebieden – bijvoorbeeld onderwijs of werving en selectie – evengoed met meer risico's gepaard kunnen gaan.[135] Ook is het belangrijk om genomen risicobeheermaatregelen te registeren. Denk aan impact assessments zoals een DPIA, audits die tijdens de levenscyclus van het algoritme zijn verricht, procedures om meldingen en klachten over het algoritme te kunnen maken en, als toepasselijk, een gemaakte risicoafweging: de motivering om het algoritme te blijven gebruiken ondanks (resterende) risico's.

Voorbeeld 3

Objectieve rechtvaardiging opnemen in het algoritmeregister

Een belangrijk risico bij de inzet van algoritmes is dat deze kunnen leiden tot bias, uitsluiting en zelfs discriminatie. Voorop staat hierbij dat organisaties de risico's hierop zo veel mogelijk moeten beperken. Als de risico's niet volledig uitgesloten kunnen worden en er goede redenen zijn om het algoritme toch te gebruiken, dan kan het gebruik van een potentieel discriminerend algoritme toelaatbaar zijn. Maar alleen in bepaalde situaties als organisaties daar een zogenaamde objectieve rechtvaardiging voor kunnen geven. Organisaties die zich in deze situatie bevinden, doen er goed aan om de motivering voor deze juridische toets in het algoritmeregister op te nemen. Juist bij afwegingen die gevoelig liggen is dit nuttig voor de legitimiteit van een algoritme. Hierdoor ontstaat inzicht in de gemaakte afwegingen en is controle, toezicht en dialoog mogelijk.[136]

Handvat 6: Faciliteer externe controle door zoveel als mogelijk transparant te zijn. Er kunnen redenen zijn om bepaalde informatie niet openbaar prijs te geven. Overwegingen die hierbij vaak genoemd worden zijn cybersecurity, bedrijfsgeheimen, mogelijkheden voor misbruik (*gaming the system*) of veiligheidsoverwegingen in bepaalde sectoren. Het uitgangspunt hierbij is dat registratie en transparantie zo veel mogelijk moeten worden nagestreefd, ondanks deze redenen. Zo kan gevoelige informatie alsnog in een niet-openbaar gedeelte van het register worden opgenomen dat bijvoorbeeld alleen toegankelijk is voor toezichthouders. Hier is ook een analogie met het regime in de AI-verordening voor AI-systemen voor rechtshandhaving, migratie, asiel en grenstoezicht. Deze zullen in een niet-openbaar deel van het register van de Europese Commissie geregistreerd moeten worden (artikel 49 lid 4 AI-verordening).

Daarnaast is transparantie niet zwart-wit en kan het in meer of mindere mate gegeven worden afhankelijk van omstandigheden. Zo kan besloten worden slechts bepaalde informatie weg te laten (zie oordeel Rechtbank). Transparant zijn geldt in het bijzonder voor overheidsorganisaties en semioverheid. Toegang tot informatie is een belangrijk democratisch principe ter verantwoording van de overheid; niet-registratie dient derhalve gemotiveerd te kunnen worden.

Voorbeeld 4

Algoritmeregister draagt bij aan nakomen transparantieverplichting oordeelt rechtbank

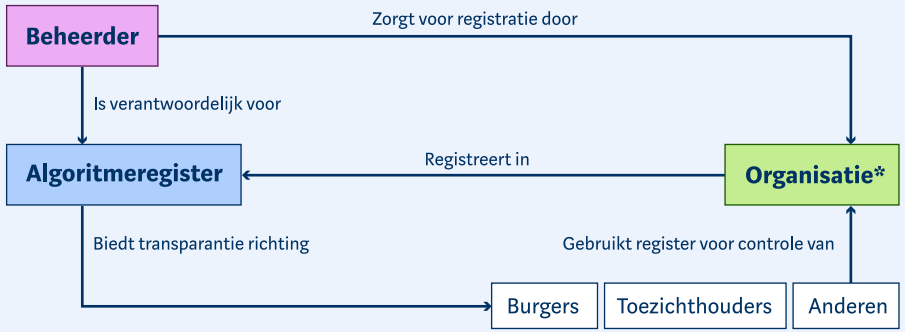
Een recente rechtelijke uitspraak onderschrijft het belang van algoritmeregistratie en het belang om zoveel als mogelijk transparant te zijn over algoritmes. Het geschil ging over informatie over algoritmes die de Belastingdienst onleesbaar had gemaakt bij een Woo-verzoek (Wet open overheid) van een journalist. De weggelaten passages zouden informatie prijsgeven over de controletechniek, waar volgens de Belastingdienst misbruik van gemaakt kan worden om controle te omzeilen. De rechtbank oordeelde dat het belang van effectieve inspectie voor de weggelaten informatie in deze situatie zwaarder woog dan het belang van de journalist om de algoritmes te controleren op discriminatie. De rechtbank betrok in de afweging dat de Belastingdienst informatie zoveel als mogelijk openbaar had gemaakt en ook was overgegaan tot het plaatsen van informatie over algoritmes in het Algoritmeregister van de Nederlandse overheid. [137]

Handvat 7: Investeer in toegankelijke vormgeving voor het register. Vormgeving speelt een belangrijke rol bij het behalen van de doelstellingen van algoritmeregistratie. Als de nadruk ligt op externe controle door veel informatie op te nemen in het register, dan is het vooral nuttig om functionaliteiten in het register in te bouwen om makkelijk informatie te kunnen zoeken of filteren op basis van verschillende categorieën (zie bijvoorbeeld het Nederlandse Algoritmeregister in box 1). Als het register verschillende doelgroepen bedient, is het ook raadzaam om informatie op meer niveaus in te delen, zoals een eerste niveau met laagdrempelige informatie (bij voorkeur op B1-taalniveau) voor burgers of klanten, waarna kan worden doorgeklikt naar meer gedetailleerde of technische informatie voor experts. Zorg ook dat het algoritmeregister goed vindbaar is, bijvoorbeeld op een duidelijke plek op de website van de organisatie. En link ernaar bij artikelen of diensten waarbij algoritmes een rol spelen. Bij registers die intern blijven kan dit bijvoorbeeld op het intranet, een ander intern kennisplatform of een gedeelde map of bestand.

Handvat 8: Zorg voor vast beheer van het algoritme-register en een duidelijke rolverdeling. Goed beheer van een algoritmeregister is noodzakelijk om zorg te dragen voor structurele activiteiten zoals functionaliteit, open-baarheid en veiligheid van het register. Daarnaast kan een beheerder van het algoritmeregister een kwaliteitsrol hebben om te waarborgen dat organisaties daadwerkelijk voldoen aan registratie-vereisten. Het beheren van het algoritmeregister is een taak en niet per se een functie die bij een specifiek persoon is belegd. De rol van beheerder zal in veel gevallen niet dezelfde zijn als de registerende rol: het organisatieonderdeel of de persoon die verant-woordelijk is dat algoritmes worden geregistreerd. Deze registrerende verantwoordelijk kan bij kleinere registers wel samenvallen met de beheerverantwoordelijkheid. Hetis daarbij belangrijk dat de beheerder voldoende mandaat krijgt om te vragen om extra informatie en anderen aan te spreken op de kwaliteit van registraties. De beheerder speelt daarnaast de rol als vraagbaak en aanspreekpunt over algoritmeregistratie intern en extern. Vanuit de registrerende verantwoordelijkheid is het belangrijk dat er binnen de organisatie een mandaat is op basis waarvan registratietaken geïntegreerd worden in verschillende bedrijfsprocessen, zoals zorgen dat bij de inkoop van nieuwe systemen in het contract wordt vastgelegd dat aanbieders van software de juiste informatie beschikbaar stellen. Figuur 3 schetst een conceptuele rolverdeling van algoritmeregistratie.

In het geval van een gedeeld algoritmeregister met algoritmes van meerdere organisaties, hangt de vraag wie de beheerder is af van de specifieke context van een sector of toepassingsgebied waarin algoritmes ingezet worden. Daarbij is het doel en scope van de algoritme-registratie (handvatten 1 en 2) van belang. Neem algoritmes in de onderwijssector: De registratie van deze algoritmes is relevant omdat deze systemen en processen impact kunnen hebben op scholieren en studenten. Hier is een situatie denkbaar dat een algoritmeregister wordt beheerd door een sectorale organisatie, bijvoorbeeld voor een bepaald onderwijsterrein (primair onderwijs, hoger onderwijs) of het Nederlandse onderwijs als geheel. Het beheer van een algoritmeregister kan bijvoorbeeld liggen bij een publieke organisatie, een brancheorganisatie of een samenwerkingsverband binnen de sector. Voor ander toepassingsgebieden, bijvoorbeeld algoritmes en AI-systemen voor HR-toepassingen of biometrische toepassingen, kan weer gelden dat in potentie alle organisaties in Nederland dit gebruiken. Voor zulke toepassingen zal een algoritmeregister weer op een andere manier beheerd moeten worden. Maatwerk is belangrijk.

FIGUUR 3 | CONCEPTUELE ROLVERDELING ALGORITMEREGISTRATIE



* Binnen de organisatie moet ook de registrerende verantwoordelijkheid belegd worden: wie (cq. Welk organisatie-onderdeel) heeft als verantwoordelijk dat algoritmeregistratie door de organisatie volledig, tijdig en correct is?

Box 2

Visie op het belang van algoritmeregistratie voor toezichthouders

Een algoritmeregister is de basis voor effectief toezicht en risicomonitoring. In deze box wordt de rol van algoritmeregistratie vanuit het perspectief van de toezichthouder beschreven.

Toezicht krijgt onder andere vorm door externe controle. Het faciliteren hiervan is een van de twee doelstellingen van algoritmeregistratie. Toezichthouders kunnen dankzij algoritmeregisters zien welke algoritmes en AI-systemen gebruikt worden en op basis van welke afwegingen. Dit creëert een grote efficiëntieslag: toezichthouders hoeven niet 'op zoek te gaan' naar algoritmes, maar kunnen zich baseren op een proactieve informatieverstrekking. Het draagt bij aan de verantwoording van algoritmes richting de toezichthouder en zorgt dat deze op tijd (nieuwe) ontwikkelingen in het vizier hebben. Dit is essentieel voor risicogebaseerd toezicht.

Voor toezichthouders is het belangrijk dat er een beheerder is aangewezen voor het proces van algoritmeregistratie. De beheerder speelt een belangrijke rol bij het algoritmetoezicht (zie figuur 3). Als er meer informatie nodig is, fungeert deze als aanspreekpunt voor de toezichthouder en andere externen.

Het is wenselijk om toezichthouders toegang te geven tot een diepere laag van het algoritmeregister... Voor het toezicht kan meer gedetailleerde informatie nodig zijn dan wat passend is voor het brede publiek. Dit helpt in de informatiehuishouding en zorgt dat de focus van toezicht kan liggen op daadwerkelijke controle van algoritmes in plaats van het zoeken naar en opvragen van informatie over algoritmes. Een register biedt een concrete bron die de toezichthouder kan gebruiken. Een algoritmeregister met meerdere lagen sluit daarnaast aan bij die situaties waarbij bepaalde informatie over algoritmes vertrouwelijk kan of moet blijven (handvat 6).

Toelichting rapportage

Deze rapportage gaat over systemen en toepassingen van algoritmes en artificiële intelligentie (AI) die impact kunnen hebben op (groepen) personen.

AI-systemen automatiseren, in de kern, handelingen en beslissingen die mensen voorheen deden. Of die voorheen niet op deze manier mogelijk waren. In simpele bewoording spreken we dan over algoritmes en AI. Dit strekt van relatief simpele toepassingen, waarin een enkel algoritme functioneert op basis van statische beslisregels, tot zeer complexe toepassingen van machine learning of neurale netwerken. De risicoanalyse in deze rapportage maakt geen onderscheid op basis van de technische werking van algoritmes en AI, waarmee wordt aangesloten bij de beleidsconsensus die ontstaat over de betekenis van de term AI-systeem.
De directie Coördinatie Algoritmes (DCA) van de Autoriteit Persoonsgegevens (AP) monitort vanuit de coördinerende AI- en algoritmetaak van de AP de mogelijke effecten van de inzet van algoritmes en AI voor publieke waarden en grondrechten. En rapporteert daar periodiek over. Dit draagt bij aan een verantwoordere inzet van AI en algoritmes.

De Rapportage AI-& Algoritmes Nederland (RAN) beschrijft trends en ontwikkelingen. De RAN beschrijft risico's die bij de inzet van algoritmes en AI die individuele personen, groepen personen of de samenleving als geheel kunnen raken. En daarmee uiteindelijk ook de samenleving kunnen ontwrichten. De AP stelt de RAN op om belanghebbenden – private en publieke organisaties, politiek, beleidsmakers en het publiek – tijdig bewust te maken van deze risico's, zodat zij actie kunnen ondernemen. Bij de beschrijving van trends en ontwikkelingen in de risico's gelden twee kanttekeningen. Ten eerste brengt de inzet van algoritmes en AI niet alleen risico's mee, maar kan deze ook positieve bijdragen leveren, ook om publieke waarden en grondrechten. In het toezicht ligt de nadruk op (het wegnemen van) risico's. Ten tweede ligt de nadruk in deze periodieke rapportage op trends en ontwikkelingen. Dit betekent dat accenten worden gelegd in de analyse, in aanvulling op structurele risico's.

De RAN bevat geen voorspellingen. De AP wil met de huidige kennis en beschikbare informatie een compact en begrijpelijk beeld geven van de huidige risico's van de inzet van algoritmes en AI en de uitdagingen bij de beheersing van deze risico's. Waar mogelijk doet de AP voorstellen voor beleid dat risico's kan tegengaan. Dit moet daarmee nog niet worden gezien als concrete guidance. De analyses en aanbevelingen in de RAN bieden organisaties en beleidsmakers inzichten om bij de inzet van algoritmes de kans op ongewenste effecten te verkleinen. Ook is de RAN te gebruiken om algoritmes en AI beter te begrijpen en de dialoog te versterken over kansen en risico's van algoritmes in de samenleving.

Dit is de vijfde editie van de RAN, die halfjaarlijks verschijnt. De inhoud is gebaseerd op de kennis die is verkregen via het toezichtnetwerk van de AP. Zoals bureau-analyse en gesprekken met meer dan honderd relevante nationale en internationale organisaties.
Maar de ontwikkelingen gaan snel en het zicht is op veel fronten nog onvolledig. Met dit in het achterhoofd probeert de AP toch een zo goed mogelijk beeld te vormen van actuele risico's en ontwikkelingen in beheersingsmaatregelen. En hieraan op een constructieve manier beleidsaanbevelingen te koppelen. Fouten of omissies in deze RAN zijn echter mogelijk.

Kom met ons in contact. Uw reacties op de RAN en suggesties zijn welkom. U kunt die mailen naar dca@autoriteitpersoonsgegevens.nl

[1] Regeerprogramma kabinet Schoof (13 september 2024). https://www.rijksoverheid.nl/documenten/publicaties/2024/09/13/regeerprogramma-kabinet-schoof

[2] BNR webredactie en ANP (28 mei 2025). Heinen en Agema zijn akkoord: 400 miljoen euro voor innovatie zorg. https://www.bnr.nl/nieuws/nieuws-politiek/10575003/akkoord-in-kabinet-over-financiering-aanvullend-zorgakkoord

[3] VZVZ (20 februari 2025). AI in de praktijk: pilot in vijf Nederlandse ziekenhuizen van start gegaan. https://www.vzvz.nl/nieuws/ai-de-praktijk-pilot-vijf-nederlandse-ziekenhuizen-van-start-gegaan

[4] UMCG (13 november 2023) UMCG beantwoordt vragen patiënten met hulp van AI

[5] IGJ-publicatie (10 februari 2025). IGJ roept zorgaanbieders op: ga zorgvuldig om met invoering van generatieve AI-toepassingen | Publicatie | Inspectie Gezondheidszorg en Jeugd

[6] Schut, M. C., Luik, T. T., Vagliano, I., Rios, M., Helsper, C. W., van Asselt, K. M., ... & van Weert, H. C. (2025). Artificial intelligence for early detection of lung cancer in GPs clinical notes: a retrospective observational cohort study. *The British Journal of General Practice, 75*(754), e316.

[7] Raad Volksgezondheid en Samenleving (15 april 2025). Iedereen bijna ziek - Over de keerzijden van diagnose-expansie. https://www.raadrvs.nl/adviezen/iedereen-bijna-ziek

[8] Corperatiemedia (13 maart 2025). De toekomst van woonruimtebemiddeling: hoe AI-agents woning-zoekenden slimmer gaan bedienen. https://www.corporatiegids.nl/nl/nieuws/de_toekomst_van_woonruimtebemiddeling_hoe_ai-agents_woningzoekenden_slimmer_gaan_bedienen-9688

[9] Redactie CorporatieNL (20 januari 2025). https://www.corporatienl.nl/artikelen/welbions-start-met-visie-op-ai-beleid/

[10] IEA (2025), Energy and AI, IEA, Paris https://www.iea.org/reports/energy-and-ai

[11] ACM (17 juli 2024).ACM marktstudie: algoritmische handel op de groothandelsmarkt voor energie https://www.acm.nl/nl/publicaties/acm-marktstudie-algoritmische-handel-op-de-groothandelsmarkt-voor-energie

[12] Niet, I. & Van Est, R. (2025) Social costs of AI in the electricity sector: Keep a close eye on sustainability and balance of power. Oxford Energy Forum (OEF) 145: 29-31 https://www.rathenau.nl/nl/klimaat/ai-de-elektriciteitssector-bewaak-duurzaamheid-en-machtsbalans

[13] E. de Winkel, Z. Lukszo, M. Neerincx and R. Dobbe. (2024). "A Review of Fairness Conceptualizations in Electrical Distribution Grid Congestion Management," IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), pp. 1-5. A Review of Fairness Conceptualizations in Electrical Distribution Grid Congestion Management | IEEE Conference Publication | IEEE Xplore

[14] European Parliament (8 juli 2022). EU strategic autonomy 2013-2023: From concept to capacity. Briefing. EU strategic autonomy 2013-2023: From concept to capacity | Think Tank | European Parliament

[15] Heath, A and Field, H. (13 juni 2025). The Verge. Meta is paying $14 billion to catch up in the AI race https://www.theverge.com/meta/685711/meta-scale-ai-ceo-alexandr-wang

[16] Algemene Rekenkamer (15 januari 2025). Publicatie Het Rijk in de cloud. https://www.rekenkamer.nl/publicaties/rapporten/2025/01/15/het-rijk-in-de-cloud

[17] Zie hoofdstuk 2 Beleid en regelgeving.

[18] RTV Noord (13 mei 2025). Noord-Drenthe en Groningen: 60 miljoen voor AI Fabriek. https://www.rtvdrenthe.nl/nieuws/17468127/noord-drenthe-en-groningen-60-miljoen-voor-ai-fabriek

[19] Kamerbrief over voorstel AI-fabriek Groningen. (27 juni 2025). https://www.rijksoverheid.nl/documenten/kamerstukken/2025/06/27/indiening-voorstel-ai-fabriek-groningen

[20] Ipsos (2025). Publication AI Monitor 2025 https://www.ipsos.com/sites/default/files/ct/publication/documents/2025-06/Ipsos-AI-Monitor-2025.pdf

[21] OECD.AI. (juni 2025). Meta AI App's Public Feed Exposes Users' Sensitive Data - OECD.AI

[22] OECD.AI. (29 mei 2025). Google Maps AI Error Causes Traffic Chaos - OECD.AI

[23] LLM Leaderboard. https://llm-stats.com

[24] De context window size van de eerste ChatGPT (op basis van GPT-3) was 1.024 karakters, terwijl de nieuwste versie (op basis van GPT-4.1) overweg kan met 1 miljoen karakters.

[25] Zo behaalt het o3-mini model van OpenAI vergelijkbare resultaten als het eerdere o1 model, tegen 15x lagere kosten.

[26] R. Bommasani, S.R. Singer, et all (17 juni 2025). The California Report on Frontier AI Policy. The Joint California Policy Working Group on AI Frontier Models.

[27] https://budget.house.gov/imo/media/doc/one_big_beautiful_bill_act_-_full_bill_text.pdf

[28] Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your Brain

on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. *arXiv preprint arXiv:2506.08872*.

29 Zie art. 37 Handvest van de EU. Zie tevens VN Mensenrechtraad, resolutie: A universal right to a healthy environment

30 EPRI (2024). White Paper. Powering Intelligence. Analyzing Artificial Intelligence and Data Center Energy Consumption.

31 World Economic Forum (2024). Fostering Effective Energy Transition WEF_Fostering_Effective_Energy_Transition_2024.pdf

32 Stanford University (april 2025). The AI Index 2025 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI. The 2025 AI Index Report | Stanford HAI

33 Stanford University (april 2025). The AI Index 2025 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI. The 2025 AI Index Report | Stanford HAI

34 Verma, P., Tan, S. (18 sepetember 2024). The Washington Post. A bottle of water per email: the hidden environmental costs of using AI chatbots. https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/

35 AI and Sustainability: Opportunities, Challenges, and Impact | EY - Netherlands

36 NOS. (22 mei 2025). Nieuw onderzoek: AI verbruikt 11 tot 20 procent van wereldwijde stroom datacenters. Nieuw onderzoek: AI verbruikt 11 tot 20 procent van wereldwijde stroom datacenters

37 Bloomberg Technology. (8mei 2025). AI is draining water from areas that needs it most bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/

38 Bolon-Canedo, V. (et all). (28 september 2024). A review of green artificial intelligence: Towards a more sustainable future. https://www.sciencedirect.com/science/article/pii/S0925231224008671

39 Wetenschappelijk Adviesraad Politie (juni 2025), Navigeren in Niemandsland: Zeven urgente uitdagingen rondom digitalisering en AI in politiewerk. https://www.wetenschappelijkeadviesraadpolitie.nl/uploads/publications/Navigeren-in-niemandsland.pdf

40 Wetenschappelijke Adviesraad Politie. (5 juni 2025). Publicatie Navigeren in niemandsland. https://www.wetenschappelijkeadviesraadpolitie.nl/uploads/publications/Navigeren-in-niemandsland.pdf

41 Ipsos (2025), AI monitor 2025.

42 Europese Commissie (2023), Discrimination in the European Union , Special Eurobarometer 535 (April-May 2023).

43 Adviescommissie uitvoering toeslagen (maart 2020), Eindadvies Omzien in verwondering 2.

44 Autoriteit Persoonsgegevens (2020), Onderzoek Belastingdienst kinderopvangtoeslag. Onderzoek Belastingdienst kinderopvangtoeslag | Autoriteit Persoonsgegevens

45 Zie hierover ook de publicatie van de Autoriteit Persoonsgegevens over betekenisvolle menselijke tussenkomst (2025). Consultatie betekenisvolle menselijke tussenkomst bij algoritmische besluitvorming | Autoriteit Persoonsgegevens

46 Hemmer, P., Schemmer, M., Vössing, M. en Kühl, N. (2021), Human AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. *PACIS 2021 Proceedings*. 78.

47 Zie hierover ook de publicatie van de Autoriteit Persoonsgegevens over betekenisvolle menselijke tussenkomst (2025).

48 College voor de Rechten van de Mens. (18 februari 2025). Meta Platforms Ireland Ltd. maakt verboden onderscheid op grond van geslacht bij het tonen van advertenties voor vacatures aan gebruikers van Facebook in Nederland. | College voor de Rechten van de Mens

49 Rechtbank Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865

50 Zie ECLI:EU:C:2025:117 (Arrest van het Hof (Eerste kamer) van 27 februari 2025, Zaak C-203/22)

51 Expertisecentrum Europees Recht (11 maart 2025). EU-Hof verduidelijkt welke informatie een verwerkingsverantwoordelijke in de context van geautomatiseerd besluitvorming moet verstrekken. https://ecer.minbuza.nl/-/eu-hof-verduidelijkt-welke-informatie-een-verwerkingsverantwoordelijke-in-de-context-van-geautomatiseerde-besluitvorming-moet-verstrekken.

52 Zie ook Autoriteit Persoonsgegevens (10 oktober 2024). Advies artikel 22 AVG en geautomatiseerd selectie-instrumenten. https://zoek.officielebekendmakingen.nl/blg-1168066.pdf

53 Zie artikel 15, 1e lid, onderdeel h van de AVG (2016/679).

54 Zie artikel 86 van de AI-verordening (2024/1689).

55 Dommering, E. (December 2023). Artificiële Intelligentie: waar is de werkelijkheid gebleven? *Computerrecht 2023/258*. https://www.ivir.nl/publicaties/download/AI-Computerrecht-2023.pdf

56 The AI Index 2025 Annual Report (april 2025). AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. The AI Index 2025 Annual Report (p. 251); MeriTalk. (29 November 2024). US Ahead of China in AI Innovation, Stanford Ranking Says

[57] Apnews (11 Februari 2025). JD Vance rails against 61 excessive AI regulation in a rebuke to Europe at the Paris AI summit

[58] The White House (23 Januari 2025). Removing Barriers to American Leadership in Artificial Intelligence

[59] Brookings (8 mei 2025). New OMB memos signal continuity in federal AI policy, https://www.brookings.edu/articles/new-omb-memos-signal-continuity-in-federal-ai-policy/;

[60] The White House (3 april 2025). M-25-21 Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. Accelerating Federal Use of AI through Innovation, Governance, and Public Trust

[61] Bird & Bird (23 Januari 2025). China TMT: Annual Review of 2024 and Outlook for 2025 (I)

[62] The AI Index 2025 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025. The AI Index 2025 Annual Report (p. 251).

[63] Europese Commissie (September 2024). The future of European competitiveness: Report by Mario Draghi

[64] D9+ Ministerial Declaration 27 Maart 2025.

[65] Europese Commissie (9 april 2025). AI Continent Action Plan COM(2025)165.

[66] Ministerie van Binnenlandse Zaken & Koninkrijksrelaties (2025). Kamerbrief uitkomsten AI-faciliteit in Nederland

[67] TechPolicy.press (24 april 2025). What's Behind Europe's Push to Simplify Tech Regulation? https://www.techpolicy.press/whats-behind-europes-push-to-simplify-tech-regulation/

[68] Ministerie van Binnenlandse Zaken & Koninkrijksrelaties (2025). Overheidsbrede handreiking voor de verantwoorde inzet van generatieve AI

[69] Autoriteit Persoonsgegevens (2023). Rapportage AI & Algoritmerisico, hoofdstuk 1 van Rapportage AI- & Algoritmerisico's Nederland (RAN) - najaar 2023| Autoriteit Persoonsgegevens

[70] Ministerie van Binnenlandse Zaken & Koninkrijksrelaties (2025). Verzamelbrief digitalisering mei 2025

[71] Ministerie van Binnenlandse Zaken & Koninkrijksrelaties (2025). Kamerbrief over voortgang motie wetenschappelijke standaard voor modellen en algoritmes

[72] Kamerstukken II 2024/2025, 32 761, nr. 322. Motie van het lid Van Nispen over algoritmes die mogelijk gebruikmaken van risicoprofilering en geautomatiseerde selectie-instrumenten in het Algoritmeregister publiceren

[73] Europese Commissie (15 mei 2025). Commission preliminarily finds TikTok's ad repository in breach of the Digital Services Act

[74] Europese Commissie (2025). Commission seeks feedback on the guidelines on protection of minors online under the Digital Services Act

[75] Kamerstukken 2024/2025, 36531, nr 12 (Amendement van het lid Six Dijkstra C.S. ter vervanging van dat gebruikt onder nr. 8).

[76] Europese Commissie (2025). De Commissie publiceert de richtsnoeren inzake verboden praktijken op het gebied van artificiële intelligentie (AI), zoals gedefinieerd in de AI-verordening

[77] Europese Commissie (2025) De Commissie publiceert richtsnoeren voor de definitie van AI-systemen om de toepassing van de regels van de eerste AI-verordening te vergemakkelijken

[78] Autoriteit Persoonsgegevens (2025). Rapportage AI & Algoritmerisico, hoofdstuk 3 van Rapportage AI- & Algoritmerisico s Nederland (RAN) Winter 2024/2025| Autoriteit Persoonsgegevens

[79] Europese Commissie (2025). AI Pact Events

[80] Europese Commissie (2025) EU Funding & Tender Portal. Tender: External service desk to provide support in complying with the AI Act

[81] Autoriteit Persoonsgegevens (2025). Oproep tot input 61 Samen werken aan AI-geletterdheid

[82] European Commission European Public Buyers Community (2025). Updated EU AI model Contractual Clauses. Updated EU AI model contractual clauses | Public Buyers Community

[83] Autoriteit Persoonsgegevens (2025). Vormvoorstel Nederlandse regulatory sandbox

[84] Crawford, K. (2021). Artificial Intelligence is Misreading Human Emotion. *The Atlantic*. https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

[85] Killoran, J., Cui, Y., Park, A., van Esch, P., & Kietzmann, J. (2023). Can behavioural biometrics make everyone happy? *Business Horizons, 66*, 585-591

[86] Stockwell, S., Hughes, M., Ashurst, C., Ní Loideáin, N. (2024). *The Future of Biometric Technology for Policing and Law Enforcement*. CeTaS Research Report. https://cetas.turing.ac.uk/publications/future-biometric-technology-policing-and-law-enforcement

[87] North-Samardzic, A. (2020). Biometric technology and ethics: Beyond security applications. *Journal of Business Ethics, 167*(3), 433-450.

[88] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124 129. https://doi.org/10.1037/h0030377

[89] Killoran, J., Cui, Y., Park, A., van Esch, P., & Kietzmann, J. (2023). Can behavioural biometrics make everyone happy? *Business Horizons, 66*, 585-591.

[90] Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (2017). Affective Computing and Sentiment Analysis. In: Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (eds) *A Practical Guide to Sentiment Analysis. Socio-Affective Computing, vol 5*. Springer, Cham. https://doi.org/10.1007/978-3-319-55394-8

[91] Artikel 3, lid 34 en 39, Artikel 5, & overweging 44, van de AI-verordening.

[92] Katirai, A. (2023). Ethical Considerations in emotion recognition technologies: a review of the literature. *AI and Ethics, 4*: 927-948.

[93] Feldman Barrett, L. (2018). *How Emotions are Made: The Secret Life of the Brain*. Mariner Books

[94] Van Heijst, K., Ploeger, A., & Kret, M. (2025). Beyond right and wrong: fostering connection in emotion theory debates. *Perspectives on Psychological Science*.

[95] Mattioli, M., & Cabitza, F. (2024). Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology. *Machine Learning and Knowledge Extraction, 6*(4), 2201-2231.

[96] Gorvett, Z. (2017, 10 april). There are 19 types of smile but only six are for happiness. *BBC*. https://www.bbc.com/future/article/20170407-why-all-smiles-are-not-the-same

[97] Zie ook overweging 44 van de AI-verordening.

[98] Wang et al., 62 A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances (2022) 83-84 *Information Fusion 19*.

[99] Clark, E.A., Kessinger, J., Duncan, S.E., Bell, M.A., Lahne, J., Gallagher, D.L. and O Keefe, S.F. (2020). The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review. *Frontiers in Psychology, 11*:920. doi: 10.3389/fpsyg.2020.00920.

[100] Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014 2023) and research recommendations. *Information fusion, 102*, 102019.

[101] D'Mello, S., & Calvo, R. A. (2013). Beyond the basic emotions: what should affective computing compute? *In CHI'13 extended abstracts on human factors in computing systems* (pp. 2287-2294).

[102] Er zijn ook indelingen die meer dimensies gebruiken, bijvoorbeeld ook dominance , zie bijvoorbeeld: Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014 2023) and research recommendations. *Information fusion, 102*, 102019.

[103] Nomiya, H., Shimokawa, K., Namba, S., Osumi, M., & Sato, W. (2025). An Artificial Intelligence Model for Sensing Affective Valence and Arousal from Facial Images. *Sensors, 25*(4), 1188.

[104] Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM, 61*(5), 90-99.

[105] Zhang, Z., Peng, L., Pang, T., Han, J., Zhao, H., & Schuller, B. W. (2024). Refashioning emotion recognition modelling: The advent of generalised large models. *IEEE Transactions on Computational Social Systems*.

[106] Zhang, Z., Peng, L., Pang, T., Han, J., Zhao, H., & Schuller, B. W. (2024). Refashioning emotion recognition modelling: The advent of generalised large models. *IEEE Transactions on Computational Social Systems*.

[107] Elyoseph, Z., Refoua, E., Asraf, K., Lvovsky, M., Shimoni, Y., & Hadar-Shoval, D. (2024). Capacity of generative AI to interpret human emotions from visual and textual data: pilot evaluation study. *JMIR Mental Health, 11*, e54369.

[108] Wake, N., Kanehira, A., Sasabuchi, K., Takamatsu, J., & Ikeuchi, K. (2023). Bias in emotion recognition with chatgpt. *arXiv preprint arXiv:2310.11753*.

[109] Een vergelijkbare observatie werd gedaan in: Hassanpour, A., Kowsari, Y., Shahreza, H. O., Yang, B., & Marcel, S. (2024, oktober). ChatGPT and biometrics: an assessment of face recognition, gender detection, and age estimation capabilities. In *2024 IEEE International Conference on Image Processing (ICIP)* (pp. 3224-3229). IEEE.

[110] Awais, M., Naseer, M., Khan, S., Anwer, R. M., Cholakkal, H., Shah, M., ... & Khan, F. S. (2025). Foundation Models Defining a New Era in Vision: a Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[111] McStay, A. (2020). Emotional AI and EdTech: serving the public good? *Learning, Media and Technology, 45*(3), 270-83.

[112] Zie bijvoorbeeld : Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research, 81*:1-15.

[113] Rhue, L. (2019, 3 Januari). Emotion-reading tech fails the racial bias test. *The Conversation*. https://theconversation.com/emotion-reading-tech-fails-the-racial-bias-test-108404

[114] Whittaker, M., Alper, M., Bennett, C.L., Hendren, S., Kaziunas, E., Mills, M., Morris, M.R., Rankin, J.L., Rogers, E., Salas, M., & Myers West, S. (2019). Disability, Bias & AI Report. *AI Now Institute*.

[115] Verhoef, T., & Fosch-Villaronga, E. (2023, September). Towards affective computing that works for everyone. *In 2023 11th International Conference on Affective*

Computing and Intelligent Interaction (ACII) (pp. 1-8). IEEE.

[116] Leijten, E. L., & Lodder, A. R. (2025). On AI That Knows How We Feel, Without Knowing Who We Are: EU Law and the Processing of Soft Biometric Data by Emotional AI. In R. Ballardini, R. van den Hoven van Genderen, & S. Järvinen (Eds.), *Emotional Data Applications and Regulation of Artificial Intelligence in Society* (pp. 93-112). (Law, Governance and Technology Series; Vol. 69). Springer Nature. https://doi.org/10.1007/978-3-031-80111-2_6

[117] Janssen, A., Kool, L., & Timmer, J. (2015). Dicht op de huid: gezichts- en emotieherkenning in Nederland. *Rathenau Instituut*.

[118] Sarah R. Blackstone, S.R., & Herrmann, L.K. (2020). Fitness Wearables and Exercise Dependence in College Women: Considerations for University Health Education Specialists. *American Journal of Health Education*, 51(4), 225-233.

[119] Piwek L., Ellis, D.A., Andrews, S., Joinson, A. (2016). The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med, 13*(2): e1001953. doi:10.1371/journal.pmed.1001953.

[120] Häuselmann, A., Sears, A. M., Zard, L., & Fosch-Villaronga, E. (2023, September). EU law and emotion data. *In 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1-8). IEEE.

[121] Overweging 44 van de AI-Verordening

[122] Handreiking AP: Aan de slag met AI-geletterdheid (2024)

[123] Zie ook de Richtlijn gegevensbescherming bij rechtshandhaving (2016/680).

[124] Zie verder ook het juridisch kader gezichtsherkenning van de AP (mei 2024), en de relevante EDPB richtsnoeren, o.a. EDPD, 63 Richtsnoeren 3/2019 inzake verwerking van persoonsgegevens door middel van videoapparatuur en EDPB, 63 Richtsnoeren 05/2022 voor het gebruik van gezichtsherkenningstechnologie in het kader van rechtshandhaving.

[125] AI-systemen voor emotieherkenning op de werkplek of in het onderwijs: Samenvatting reacties van reacties en vervolgstappen (AP, februari 2025)

[126] Zo heeft de EDPB bijvoorbeeld in december 2024 aangekondigd richtsnoeren te gaan publiceren over de wisselwerking tussen de AVG en de AI-verordening.

[127] Ovw. 14 AI-verordening.

[128] Artikel 4 (14) AVG, zie ook ovw. 51 AVG; Artikel 3 (34) AI-verordening, zie ook ovw. 14 AIV.

[129] Zie juridisch kader gezichtsherkenning, p.10-11 en artikel 9(2) AVG.

[130] Art. 5 (1) (f) AI-verordening

[131] Art. 5 (1) (g) AI-verordening

[132] Artikel 3(39) AI-verordening en Ovw. 18 AI-verordening; zie ook EC richtsnoeren inzake verboden AI-praktijken onder de AIV, para. 250-252, zie m.n. ook voetnoot 160 (p. 84)

[133] E. Nieuwenhuizen 2025, *Algorithm Registers: A Box-Ticking Exercise or Meaningful Tool for Transparency?* Information Policy, 29(4), 415-433.

[134] C. Cath, F. Jansen 2022, *Dutch Comfort: The limits of AI governance through municipal registers*. Techné: Research in Philosophy and Technology, 26:3, 395-412.

[135] Autoriteit Persoonsgegevens. Risicogroepen AI-verordening. https://www.autoriteitpersoonsgegevens.nl/themas/algoritmes-ai/ai-verordening/risicogroepen-ai-verordening

[136] Zie ook: Hoofdstuk 2, Rapportage AI- & Algoritmerisico's Nederland Februari 2025

[137] Rechtbank Den Haag 20 mei 2025, ECLI:NL:RBDHA:2025:9525.

AP | bescherming in een digitale wereld