Rapportage algoritmerisico's Nederland

Rapportage juli 2023

Autoriteit Persoonsgegevens | Directie Coördinatie Algoritmes (DCA)

Periodiek inzicht in risico's en effecten van
de inzet van algoritmes in Nederland

AUTORITEIT
PERSOONSGEGEVENS

Inhoudsopgave

AUTORITEIT
PERSOONSGEGEVENS

Introductie

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

Deze AI-systemen automatiseren, in de kern, handelingen en beslissingen die voorheen door mensen gedaan werden. Of die niet op deze wijze mogelijk waren. Eenvoudig gezegd: het gaat over algoritmes en AI. Dit strekt van relatief simpele toepassingen, waarin een enkel algoritme functioneert, tot zeer complexe toepassingen van ***machine learning*** of neurale netwerken. Voor de risicoanalyse maken we in de definitie geen onderscheid tussen algoritmes of AI. Dit rapport spreekt in alle gevallen over 'algoritmes' en 'systemen en/of toepassingen'. In beginsel betreft dit dan 'algoritmes' en 'systemen en/of toepassingen' die (groepen) personen kunnen raken. Door de risico's en effecten van de inzet van algoritmes en AI voor publieke waarden en grondrechten te monitoren en hierover periodiek te rapporteren, draagt de directie Coördinatie Algoritmes (DCA) van de Autoriteit Persoonsgegevens (AP) bij aan een verbetering van verantwoorde inzet van algoritmes. Publieke waarden en grondrechten zijn bijvoorbeeld non-discriminatie, transparantie en uitlegbaarheid, voorkomen van misleiding, vrijheid van meningsuiting en kansengelijkheid.

De Rapportage Algoritmes Nederland (RAN) beschrijft risico's. Dit betreft risico's bij de inzet van algoritmes die individuele personen, groepen personen of de samenleving als geheel kunnen raken. En daarmee uiteindelijk ook de samenleving kunnen ontwrichten. De DCA stelt de RAN op om belanghebbenden - private en publieke organisaties, politiek, beleidsmakers en het publiek - tijdig bewust te maken van deze risico's, zodat actie ondernomen kan worden. Het eerste hoofdstuk van deze rapportage beschrijft op hoofdlijnen de voor Nederland belangrijkste recente ontwikkelingen op het gebied van de inzet van algoritmes en de risicobeheersing daarvan. Het tweede hoofdstuk is thematisch en analyseert relevante praktijkvoorbeelden in meer detail. Het derde hoofdstuk schenkt aandacht aan beleidsontwikkeling en institutionele kaders.

De RAN bevat geen voorspellingen. De DCA beoogt met de huidige kennis en beschikbare informatie een compact en begrijpelijk beeld te geven van de huidige risico's en bijbehorende beheersingsuitdagingen van de inzet van algoritmes.

Waar mogelijk doet de DCA voorstellen voor beleid dat risico's kan tegengaan. De analyses en aanbevelingen in de RAN bieden organisaties en beleidsmakers inzichten om bij de inzet van algoritmes de kans te verkleinen op ongewenste effecten voor grondrechten, publieke waarden en fundamentele vrijheden. Ook is de RAN een methode om algoritmes beter te begrijpen en de dialoog te versterken over kansen en risico's van algoritmes in de samenleving.

Uw reacties op de rapportage en suggesties voor verbetering zijn welkom. U kunt die mailen naar dca@autoriteitpersoonsgegevens.nl.

Kernboodschappen

- Het is mogelijk om algoritmes op een verantwoorde manier te ontwikkelen en in te zetten, hiermee kan maatschappelijke waarde worden geboden. Maar de inzet van deze algoritmes brengt ook risico's mee die beheerst moeten worden.
- Het belang van adequate risicobeheersing gaat de komende tijd verder toenemen, omdat algoritmes krachtiger worden en nieuwe toepassingsmogelijkheden en risico's ontstaan.
- De Nederlandse samenleving moet stappen (blijven) zetten om grip te krijgen op de maatschappelijke effecten van de inzet van algoritmes en artificiële intelligentie (AI). Dat vraagt om: (i) snellere totstandkoming van (wettelijke) eisen aan transparantie, (ii) duidelijke regulering en (iii) versterking van intern en extern toezicht. Het vraagt ook dat organisaties meer personeel en scholing inzetten om algoritmerisico's te beheersen.
- Twijfelen organisaties aan de adequaatheid van hun risicobeheersing? Dan doen zij er verstandig aan terughoudend te zijn met de inzet van algoritmes en AI.
- Kunnen algoritmes en AI grote maatschappelijke impact hebben? Dan is het wenselijk dat organisaties voor, tijdens en na de inzet hiervan verantwoording afleggen. Praktijkvoorbeelden laten zien dat ook in deze gevallen organisaties algoritmes nog te vaak zien als een puur interne, organisatorische aangelegenheid.
- De AP is positief over het algoritmeregister voor overheidsorganisaties. Ook ziet de AP ruimte voor een verplichte, maar risicogebaseerde vulling van het register. Er moet een deadline komen voor eerste vulling met hoog-risico algoritmes. Overheidsalgoritmes en hun risico's zullen in de eerste helft van 2024 in kaart gebracht moeten zijn om te weten of deze als hoog risico worden geclassificeerd onder de AI-verordening. Deze algoritmes zullen dan ook in het algoritmeregister geregistreerd moeten worden. De DCA zal nadrukkelijk de voortgang binnen de overheid volgen en hierover rapporteren. Van private organisaties met een significant maatschappelijke taak mag worden verwacht dat zij proactief stappen zetten richting meer transparantie en uitlegbaarheid.

Samenvatting risicobeeld en aanbevelingen

Deze rapportage geeft een eerste beeld van de risico's en effecten van de inzet van algoritmes en artificiële intelligentie (AI) in Nederland.

De rapportage is een product van de nieuwe directie Coördinatie Algoritmes (DCA) van de Autoriteit Persoonsgegevens (AP). De DCA signaleert en analyseert risico's en effecten van de inzet van algoritmes over alle sectoren en domeinen van de samenleving heen. De DCA richt zich op algoritmes waarvan de risico's en effecten direct of indirect (groepen) personen, of alle mensen in de samenleving treffen.

Algoritmes die primair worden toegepast als een veiligheids- of sturingscomponent in een product, systeem of proces, vallen veelal buiten de reikwijdte van de overkoepelende monitoring die de DCA met de Rapportage Algoritmerisico's Nederland beoogt. Van deze algoritmes mag doorgaans verwacht worden dat deze in beginsel geen impact hebben op publieke waarden en grondrechten. Als onderdeel van haar signalerende taak publiceert de DCA elk halfjaar de RAN. Aan het eind van 2023 verschijnt de eerste volledige rapportage. Die zal mede gebaseerd zijn op risicosignalen die wij van andere toezichthouders ontvangen. Deze eerste rapportage is ook bedoeld om de rapportagemethodiek te ontwikkelen.

Verantwoorde ontwikkeling en inzet van algoritmes die (groepen) personen kunnen raken is mogelijk.

Dit vergt inspanning en gaat gepaard met risico's die actief beheerst moeten worden; voorafgaand, tijdens en na inzet. Er moet niet enkel gewerkt worden aan de ontwikkeling van algoritmes, maar ook aan een ecosysteem van risicobeheersing en verantwoording. Wanneer onvoldoende aandacht wordt besteed aan de verantwoorde ontwikkeling en inzet van algoritmes loopt Nederland mogelijk belangrijke en positieve toepassingen van algoritmes mis.

Het voorlopige eerste beeld is dat de Nederlandse samenleving stappen moet (blijven) zetten om grip te krijgen op algoritmes.

De politiek speelt hierbij – ook in de demissionaire fase van kabinet-Rutte IV – een grote rol. Het bewustzijn over en de aandacht voor algoritmerisico's neemt nadrukkelijk toe. Maar het ontbreekt over de gehele linie nog te vaak – maar

zeker niet in alle gevallen – aan volwassen, structurele en bindende algoritmespecifieke instrumenten om alle risicovolle algoritmes adequaat te monitoren en algoritmerisico's te beheersen. Het werken aan een beheersbare en gereguleerde inzet van algoritmes vraagt juist nu om politiek draagvlak, omdat algoritmische systemen en toepassingen zich snel blijven ontwikkelingen en steeds breder worden toegepast. De AP roept daarom het demissionaire kabinet, de Eerste Kamer en Tweede Kamer op om voortgang te maken of zelfs te versnellen met de inspanningen voor beleid, implementatie en intern- extern toezicht dat ziet op algoritmes.

Waar structurele en bindende kaders ontbreken, is het aan organisaties binnen overheid en bedrijfsleven zelf om een passende aanpak te bepalen.

Dit dient aanvullend op en in verbinding met naleving van bestaande wet- en regelgeving zoals de Algemene verordening gegevensbescherming (AVG) te gebeuren. Daarin worden soms al belangrijke elementen verplicht. De mate van volwassenheid en risicobeheersing verschilt per sector en organisatie. Het zicht op ontwikkelingen en toepassingen is nog onvoldoende helder voor een volledig overkoepelend beeld. De DCA werkt hier doorlopend aan en sectorale toezichthouders spelen een leidende rol in het komen tot uniforme risicobeheersing toegespitst op sectoren. De DCA verwacht dat de grootste uitdaging zit bij die sectoren, toepassingen of innovaties waar géén sectorale toezichthouder is die toeziet op het algehele functioneren van organisaties in die sector.

Initiatieven zoals registers, productstandaardisatie, beoordelingskaders en audittechnieken bevinden zich vaak nog in de pilotfase.

Ook deze eerste rapportage is daar in zekere zin een voorbeeld van. Het is pionieren om een totaalpakket aan beheersmaatregelen te ontwikkelen voor systemen die al in alle domeinen van de samenleving zijn doorgedrongen. Het is daarom belangrijk dat beleidsmakers doorpakken om instrumenten en principes als registers, toetsingskaders en transparantie verplicht te maken bij de ontwikkeling en inzet van risicovolle algoritmes. 'Het betere is daarbij de vijand van het goede': met eenvoudige regels kan al veel worden bereikt. Organisaties moeten vervolgens zorgen voor voldoende financiële en personele capaciteit om snel aan dit soort regels te voldoen. Zij moeten daarin worden ondersteund als dat nodig is.

Een inhaalslag in het opbouwen van de beheersing van risicovolle 'traditionele' algoritmes is nodig...

Veel organisaties staan nog aan het begin van meer transparantie over risicovolle maar tegelijkertijd simpele algoritmes die zij gebruiken – soms al vele jaren. De zorg betreft met name die organisaties die geen geïnstitutionaliseerde aanpak en geen duidelijke verantwoordelijkheid hebben ingericht voor het checken van mogelijke bias en eerlijkheid van algoritmes, gericht op de samenleving, groepen en individuen. Dit geldt zowel voorafgaand aan als tijdens het gebruik van algoritmes. Wat betreft de algehele staat van risicobeheersing is een kanttekening bij deze observatie op zijn plaats, want het is voor een coördinerende toezichtautoriteit als de DCA op dit moment lastig observeren. Het ontbreekt vooralsnog aan een (gestructureerd) overzicht over de wijze waarop verschillende types organisaties binnen verschillende maatschappelijke sectoren hun beheersing al dan niet hebben vormgegeven. Daarbij geldt dat meer zicht op organisaties die 'goede voorbeelden' zijn, ook kan helpen in de uitleg aan organisaties waarbij de risicobeheersing nog in opbouw is.

...maar de komst van geavanceerde en complexe algoritmes maakt die uitdaging extra groot.

Deze meer complexe algoritmes, die bijvoorbeeld gebaseerd zijn op (zelf)lerende neurale netwerken, zijn bijzonder uitdagend op het gebied van toetsing, transparantie en evaluatie. Dit vergroot de uitdaging voor organisaties om in een snel tempo een effectieve risicobeheersing te implementeren. Dit geldt des te meer omdat bij complexe algoritmes de risico's zich op een andere manier materialiseren dan bij simpele algoritmes. En de beheersingsmethodiek daarom weer anders vorm moet krijgen. Ditzelfde geldt voor transparantie, verantwoording en uitlegbaarheid. Bij complexe algoritmes, zoals neurale netwerken, is dit een specifiek vakgebied dat nog in ontwikkeling is ***(Explainable AI)***.

Zo lang er nog onvoldoende risicobeheersing is, doen organisaties er daarom verstandig aan terughoudend te zijn met de inzet van algoritmes.

De AP waarschuwt voor inzet van nieuwe toepassingen zo lang mogelijke risico's op de schending van grondrechten en publieke waarden nog niet zijn geïnventariseerd. Dit geldt voor de inzet van alle types algoritmes in de processen van organisaties, waarbij deze inzet de samenleving, groepen of individuen kan raken. Denk hierbij aan systemen en toepassingen zoals gezichtsherkenning of voor het opsporen van

frauderisico's. Dat is niet alleen juridisch verplicht, bijvoorbeeld op grond van de AVG, maar het is ook de consequentie van de lessen die organisaties zouden moeten trekken uit de ontwrichtende casuïstiek van de afgelopen jaren. Het besef is diep doorgedrongen dat actief gestuurd moet worden op het voorkomen van negatieve gevolgen, zoals discriminatie, willekeur en misleiding. Een organisatie moet daar organisatorisch klaar voor zijn, voordat de organisatie een systeem of toepassing met algoritmes ontwikkelt.

Organisaties die koplopers zijn bij de ontwikkeling en inzet van algoritmes moeten zich extra bewust zijn van de inspanning die van hen wordt gevraagd om publieke waarden en grondrechten te waarborgen.

De algoritmes die deze organisaties gebruiken, zijn vaak impactvoller. Daardoor kan bij fout gebruik het vertrouwen in algoritmes ernstig worden ondermijnd. Nieuwe algoritmische technieken zijn krachtig, maar in potentie ook ontwrichtend. Bijvoorbeeld doordat zij een valse werkelijkheid kunnen creëren. Denk aan ***deepfakes***. Hoe meer dit soort geavanceerde systemen wordt toegepast, hoe moeilijker het voor mensen wordt om zeker te weten wat wel en niet te vertrouwen is. Of zij nou burger, klant of werknemer zijn.

Nederland staat zeker niet alleen in deze situatie - internationale samenwerking is cruciaal.

Wereldwijd spelen dezelfde uitdagingen. De Europese Unie (EU) heeft de kans voorop te lopen als eind dit jaar een politiek akkoord wordt bereikt op de AI-verordening. De AP ziet dat de conceptwetgeving steeds meer aandacht schenkt aan het beschermen van grondrechten en fundamentele vrijheden. Zoals het recht om niet gediscrimineerd te worden of de vrijheid van meningsuiting. Dit is een beweging in de goede richting.

De AI-verordening is belangrijk, maar gaat tegelijkertijd geen panacee zijn voor de huidige uitdagingen.

Allereerst is het compliance-mechanisme voor een grote groep toepassingen met een hoog risico gebaseerd op de eigen beoordeling van de producent van het systeem. Er is vooraf dus geen onafhankelijke zekerheid dat een systeem geen inbreuk vormt op grondrechten. Daarnaast is de regelgeving na een akkoord pas over enkele jaren bindend. Hierop vooruitlopend moet daarom worden nagedacht over manieren om de belangrijkste bepalingen al wel betekenis te geven. Gegeven de wereldwijde ontwikkeling en inzet van systemen en toepassingen als ***large language models*** (LLM's), moet regulering ook grensoverschrijdend zijn om te kunnen profiteren van innovatie. Het is verstandig hierbij een Europese of mondiale aanpak na te streven, in lijn met de ontwikkelingen in de EU en de recente afspraken hierover door de G7.

Nadelige gevolgen van bestaande algoritmes bevinden zich vaak onder de radar, risico's van nieuwe technologieën staan direct in de spotlights.

Bij gebrek aan overzicht is het eigenlijk niet mogelijk om te bepalen welke algoritmes, die op dit moment in gebruik zijn, de grootste zorgen opleveren over publieke waarden en grondrechten. Vanuit bekende casuïstiek, bijvoorbeeld op het gebied van rechtshandhaving, betalingsverkeer of sociale voorzieningen, zijn al wel toepassingen aan te wijzen waarbij er kans is op (groeps)discriminatie, (groeps)oneerlijkheid of gebrek aan verantwoording. Het vergt systematische rapportage en beoordeling om ook vanuit het perspectief van continue risicobeoordeling hierop grip te krijgen. Mogelijke problemen bij gegevensbescherming in LLM's worden in Europa al onderzocht en waar nodig geadresseerd vanuit bestaande wet- en regelgeving. Voor de AP bestaat het kader voor mogelijke onderzoeken naar systemen en toepassingen uit de AVG en de Wet politiegegevens (Wpg).

Praktijkvoorbeelden laten zien dat maatschappelijk significante organisaties de inzet van algoritmes in hun kernprocessen soms zien als een uitvoeringskwestie en interne keuze, waarover zij voor, tijdens en na de inzet geen of beperkte verantwoording hoeven af te leggen.

Het Criminaliteits Anticipatie Systeem (CAS) van de Nederlandse politie is een voorbeeld van een algoritmisch systeem dat in de gehele Nederlandse samenleving wordt ingezet. De technische informatie die publiekelijk bekend is over dit algoritme, is beperkt en wordt niet geactualiseerd. Nederlandse financiële instellingen gebruiken algoritmes om hun wettelijke taak te ondersteunen om transacties te monitoren voor antiwitwas- en sanctiemaatregelen. Dit soort algoritmes brengt het risico met zich mee dat het kan bijdragen aan ongewenste discriminerende effecten. Financiële instellingen gebruiken daarom technische evaluatietools om hun modellen te toetsen op bias, maar ook hier ontbreekt het aan (structurele en publieke) transparantie over welke evaluatietools zij precies gebruiken. Meer transparantie en bijbehorende begrijpelijke uitlegbaarheid kan het publieksvertrouwen verhogen en kan bijdragen aan het verbeteren van de kwaliteit van deze tools.

1. Overkoepelende ontwikkelingen

Bewustzijn ongewenste effecten

De bewustwording van de ongewenste effecten van de inzet van algoritmes neemt toe, terwijl de technologie in rap tempo door ontwikkelt. In alle delen van de samenleving neemt het gebruik van algoritmes toe. En worden mensen zich steeds meer bewust van de effecten van de inzet van algoritmes. Een voorbeeld van een recente ontwikkeling is generatieve AI, bijvoorbeeld in de vorm van large language models. Veel Nederlanders zijn inmiddels bekend met ChatGPT van OpenAI. Door het verzamelen, doorzoeken en analyseren van grote hoeveelheden tekst kan een chatbot antwoorden op vragen formuleren of kan een LLM taken uitvoeren. Zoals het geautomatiseerd maken van presentaties, plannen, documentatie of brieven. Ook voor afbeeldingen en video zijn toepassingen in opkomst, die via ***text-to-image*** een LLM gebruiken om afbeeldingen of video te genereren op basis van tekst. Ook deze visuele toepassingen vinden al hun weg naar bestaande tools. Het gebruik van LLM's biedt kansen om processen bijvoorbeeld efficiënter te laten verlopen. Maar bij de inzet van deze nieuwe soort toepassing van algoritmes leven er ook steeds meer (soms nieuwe) maatschappelijke zorgen. Bijvoorbeeld over de snelle verspreiding van nepnieuws, privacyinbreuken en auteursrechtschendingen.

Het vertrouwen in algoritmes neemt af. De bekendheid met algoritmes is de afgelopen jaren toegenomen, maar het vertrouwen in algoritmes neemt vaker af dan toe (zie figuur 1). Onderzoek van KPMG, uitgevoerd door Motivaction, toont dat ook het absolute niveau van vertrouwen laag is. Slechts 20 tot 30% van de personen heeft vertrouwen in het gebruik van algoritmes. Het maakt daarbij niet uit of het om algoritmes van overheidsorganisaties gaat, financiële instellingen, retailorganisaties, zorginstellingen, boekingswebsites of techbedrijven. Volgens de ondervraagde personen kunnen organisaties dit vertrouwen verreweg het best vergroten door transparanter te zijn (ca. 40%) en algoritmes te richten op het verbeteren van de samenleving in plaats van het opsporen van incidenten (ca. 25%).

FIGUUR 1: NEDERLANDSE BURGERS EN ALGORITMES

BRON: KPMG (2022) – ONDERZOEK VERTROUWEN NEDERLANDSE BURGER IN ALGORITMES

In de afgelopen maanden waarschuwden verschillende organisaties en instanties voor diverse effecten van de inzet van generatieve AI. Prominente ontwikkelaars en AI-experts waarschuwden, mede gevoed door publieke en politieke druk, voor de snelle ontwikkeling. En stelden een ontwikkelpauze van een half jaar voor, om te focussen op de effecten voor de samenleving. Hoewel deze zorgen de moeite waard zijn om te bespreken, kunnen ze ook afleiden van de daadwerkelijke vraagstukken waarmee de samenleving nu geconfronteerd wordt. En van instrumenten om algoritmes verantwoord te ontwikkelen en in te zetten.

In Nederland zette de minister van Onderwijs, Cultuur en Wetenschap (OCW) eind juni een fraude-opsporingsalgoritme bij de Dienst Uitvoering Onderwijs (DUO) per direct stop. Dit gebeurde naar aanleiding van journalistiek onderzoek naar mogelijke discriminerende effecten van het algoritme. Het besluit om de inzet van dit algoritme per direct stop te zetten, bevestigt het beeld dat de bewustwording van algoritmerisico's hoog is. Maar ook dat de interne risico-identificering en -beheersing binnen – in dit geval – overheidsorganisaties nog niet robuust genoeg is om op te vertrouwen. Een stabiele beheersing van algoritmerisico's ontstaat pas wanneer organisaties signalen zoals deze al in een vroeg stadium ondervangen via interne evaluerende risicobeoordelings- en beheersingssystemen periodiek worden ondervangen en waar nodig geadresseerd door (de inzet van) het algoritme aan te passen.

Tegelijkertijd mogen, ook in de rapportage algoritmerisico's Nederland, de positieve effecten niet vergeten worden. In groeiende mate worden toepassingen zichtbaar die een tastbare bijdrage leveren voor individuen, groepen en de samenleving. Uiteraard zijn dit toepassingen in de gezondheidszorg en voor mensen met een beperking. Maar ook toepassingen in de industrie, landbouw, infrastructuur en het maatschappelijk middenveld laten zien dat innovatie, mits verantwoord, absoluut een positieve bijdrage kan leveren aan de samenleving.

Risicobeeld

Om complexe algoritmes beheersbaar te maken is door de samenleving als geheel aanvullende inspanning nodig… Met de razendsnelle ontwikkelingen en de grote maatschappelijke belangen die op het spel staan, moeten alle betrokken partijen extra inspanning leveren voor de beheersbaarheid. Het aantal toepassingen en mogelijkheden groeit dagelijks, zoals zichtbaar bij LLM's. Door deze snelle ontwikkelingen en niet altijd sluitende of heldere regelgeving loopt de Nederlandse samenleving het gevaar dat risicovolle algoritmes onvoldoende beheerst in de haarvaten van de samenleving terecht komen. Voor het breed inzetten van een systeemtechnologie is een passend kader nodig, zeker in het geval van generatieve toepassingen. Daarin moeten publieke waarden en grondrechten beschermd worden en in balans zijn met innovatie.

…tegelijkertijd blijft de situatie dat ook simpele algoritmes grote schade kunnen aanrichten. Simpele systemen en toepassingen waarin algoritmes maar een beperkte rol spelen, kunnen ook al tot ingrijpende gevolgen leiden voor individuen, groepen en de samenleving. Voorbeelden hiervan zijn de Toeslagenaffaire en frauderisicosystemen die door gemeenten worden ingezet. Vaak gaat het om systemen en toepassingen gebaseerd op historische data of impliciete veronderstellingen waarmee risicofactoren uit die data worden afgeleid. Hiermee wordt niet daadwerkelijk fraude of ander te signaleren gedrag vastgesteld, maar wordt enkel het risico op dat gedrag afgeleid uit data over vastgestelde gedragingen in het verleden. Niet eerder geconstateerde gedragingen, spurieuze verbanden (schijnverbanden) of onterechte aannames leiden vaak tot een vertekend beeld van de werkelijkheid. Het gevaar van discriminatie, gebrek aan transparantie of onbedoelde willekeur is hier sterk aanwezig.

Voor veel mensen blijft de inzet van algoritmes verborgen door een gebrek aan transparantie, ze weten simpelweg niet dat een algoritme in het spel is. Veelal schuurt dit gebrek aan transparantie met bestaande wet- en regelgeving. Niet naleven van bestaande wet- en regelgeving maakt de kloof nog groter ten opzichte van toekomstige aanvullende wet- en regelgeving. Voor de samenleving, groepen en vooral het individu is transparantie bij systemen en toepassingen die invloed kunnen hebben op publieke waarden en grondrechten een cruciaal onderdeel. Zonder transparantie verliezen individuen en groepen de controle, kunnen zij maar beperkt iets doen tegen besluiten en wordt het betwisten van uitkomsten en effecten praktisch onmogelijk. Wanneer het ontbreekt aan transparantie of uitleg over de inzet van een algoritme, kunnen manipulatieve toepassingen voorkomen. Beïnvloeding van (online) gedrag door profilering en kleine of grote besluiten die door algoritmes genomen worden, kunnen tot ernstige effecten leiden. Bijvoorbeeld bij kansspelen of zelfs democratische processen.

Organisaties in de voorhoede van de inzet van nieuwe technologie en nieuwe algoritmische systemen en toepassingen moeten zich bewust zijn van de extra inspanning die dit van hen vraagt. Om nuttige en verantwoorde systemen en toepassingen in de samenleving te kunnen aanbieden of inzetten, moeten risico's beheerst worden. Juist bij nieuwe inzet vraagt dit meer van organisaties dan bij bestaande en bekende technologieën. Het in kaart brengen van risico's is een lastiger proces, maar noodzakelijk omdat

veel risico's niet volledig uitgesloten kunnen worden. Daarom vraagt dit niet enkel voorafgaand aan maar ook tijdens de inzet om monitoring, beheersing en een volwassen organisatie. Zo'n organisatie voldoet aan geldende wet- en regelgeving, heeft passende kennis en kunde, ontwikkelt op een verantwoorde wijze en blijft de inzet monitoren. Ook kleine of startende ondernemingen kunnen in die zin een volwassen organisatie zijn. Zonder deze volwassen en verantwoorde opstelling zullen de samenleving nuttige, waardevolle systemen en toepassingen worden ontzegd. Juist om de kansen van algoritmes te benutten, moet nu geïnvesteerd worden in het beheersen van risico's, verantwoorde ontwikkeling en robuuste organisaties die durven te ontwikkelen binnen de democratisch vastgestelde kaders.

Toenemende inzet en afhankelijkheid van algoritmes kan hand in hand gaan met toenemende (markt)macht voor grote techbedrijven. Veel complexe algoritmes vereisen enorme rekenkracht en enorme hoeveelheden gegevens. Het risico is dat slechts een kleine groep techbedrijven de reken- en gegevenskracht heeft om de meest geavanceerde modellen te ontwikkelen. Grondrechten, democratie en de rechtstaat bieden ook essentiële bescherming tegen machtsconcentraties en het misbruiken daarvan. In de digitale samenleving zijn consolidaties van macht niet voorbehouden aan de staat, maar hebben ook private technologiepartijen zeer grote macht.

De verdere opkomst van algoritmes kan leiden tot fundamentele verschuivingen in onze samenleving. Passende educatie van burgers is van belang om hen niet enkel digitaal vaardig te maken, maar ook bewust van de werking en risico's van algoritmes in de samenleving. Maar ook bestaande stelsels kunnen een volledig nieuwe benadering vereisen door de impact van algoritmes. Dit stelt onder meer het IMF. Bijvoorbeeld het onderwijs (de arbeidsmarkt verandert), de zorg (andere wijze van diagnostiek en andere rol medisch specialisten) en de overheid (verhouding tussen belasting van kapitaal en arbeid). Op basis van internationale AI-standaarden, zoals van UNESCO, moet ook worden stilgestaan bij vraagstukken als de impact van algoritmes op mensen met beperkingen en de klimaatimpact.

Om risico's van de inzet van AI te beheersen moeten organisaties zich inspannen voor transparantie, dialoog met de samenleving en anticipatie op nieuwe wet- en regelgeving. Allereerst draagt transparantie bij de ontwikkeling en inzet van systemen en toepassingen bij aan vertrouwen in innovatie en begrip voor de werking van systemen en toepassingen in de samenleving. Maar het biedt ook perspectief voor mensen die mogelijk negatieve gevolgen ondervinden of waarbij een ***chilling effect*** optreedt. Handvatten om transparantie te bieden zijn voor veel sectoren, terreinen en toepassingen al te vinden in bestaande wet- en regelgeving. Het naleven hiervan is daarom vanzelfsprekend van groot belang. Ten tweede moeten organisaties zich bewust zijn van mogelijke risico's en effecten voor publieke waarden en grondrechten als zij systemen en toepassingen ontwikkelen en inzetten. Dit vereist een continue dialoog met de samenleving over deze risico's en effecten. Bijvoorbeeld door stakeholders te raadplegen bij de ontwikkeling en het monitoren van de risico's en effecten. Ten derde adviseert de DCA dringend om de principes van nieuwe wet- en regelgeving nu al mee te nemen in de ontwikkeling en inzet van systemen en toepassingen. Om die belangrijke stap vooruit te zetten, kan het helpen om de diverse internationale ***frameworks*** met belangrijke principes te bestuderen of hierbij aansluiting te zoeken. Bijvoorbeeld de tien AI-principes van UNESCO. Het kan bijdragen aan verantwoorde innovatie als in ieder geval elementen van nieuwe wetgeving of relevante frameworks in een organisatie aanwezig zijn. Zonder op toekomstige wetgeving te anticiperen, kunnen organisaties niet op een verantwoorde wijze tegemoet komen aan de innovatiewens van de samenleving. Daarmee kan de samenleving maatschappelijk relevante innovatie mislopen. Bijvoorbeeld op belangrijke thema's als gezondheid, klimaat en democratie.

Het chilling effect

Het ***chilling effect*** duidt op het verschijnsel dat mensen hun gedrag aanpassen wanneer ze het gevoel krijgen in hun grondrechten te worden aangetast. Ongeacht of dit echt zo is.

In een winkelstraat met zichtbaar cameratoezicht gedragen mensen zich anders, doordat zij zich bewust zijn van de camera's. Het idee dat we bekeken worden, werkt al afschrikwekkend of ontmoedigend voor ons gedrag. Dit effect geldt net zo goed als de camera's niet functioneel zijn.

Een nieuwe vorm van het chilling effect doet zich voor in de interactie tussen mens en algoritmes. Het is namelijk ook mogelijk dat mensen hun gedrag veranderen als zij weten, of het idee hebben, dat ze niet door een mens maar door een algoritme zullen worden beoordeeld. Bijvoorbeeld wanneer werkzoekenden niet meer beginnen aan een sollicitatieprocedure als ze weten dat een algoritme hen mogelijk beoordeelt. Het chilling effect verandert dan het gedrag van mensen, zonder dat er een directe schending van rechten heeft plaatsgevonden.

2. Algoritmes in de praktijk

In de afgelopen periode stonden meerdere types algoritmische systemen die in Nederland gebruikt worden in de spotlights.

Zo publiceerde het Fundamental Rights Agency (FRA) van de EU in december 2022 een rapport waarin uitgebreid werd ingegaan op het risico van (groeps)discriminatie door algoritmes voor ***predictive policing*** (PP). Daarnaast stelde de Nationaal Coördinator tegen Discriminatie en Racisme (NCDR) in april 2023 signalen te ontvangen over discriminatie door financiële instellingen. Verder heeft de AP in mei 2023 vijf gemeenten om opheldering gevraagd over het gebruik van de 'fraudescorekaart'. Dit is een algoritme om het risico op bijstandsfraude in beeld te brengen. Naar aanleiding van deze signalen biedt dit hoofdstuk een korte beschrijving van drie types systemen voor handhaving, betalingsverkeer en gemeentelijke sociale voorzieningen. De DCA beschrijft in deze rapportage de algoritmische systemen en toepassing vanuit de overkoepelende coördinerende algoritmetaak. Het betreft daarom geen beschrijving vanuit het perspectief van de AVG-toezichttaak.

Rechtshandhaving

Bias en vervolgens (groeps)discriminatie zijn grote risico's bij PP. Algoritmes worden veel ingezet bij wetshandhavingstaken, op verschillende manieren, onder andere bij PP. Het Criminaliteits Anticipatie Systeem (CAS) van de Nederlandse politie is een bekend voorbeeld. Dit is het afgelopen jaar veel besproken, bijvoorbeeld in de Tweede Kamer, door de Algemene Rekenkamer en door het FRA. PP is het gebruiken van algoritmes om misdrijven door bepaalde personen of op bepaalde locaties en tijdstippen te voorspellen. Het CAS doet dat laatste: het deelt Nederland op in vakken van 125 meter en voorspelt de kans op (de aangifte van) een misdrijf in zo'n vak. Het is wereldwijd het enige operationele PP-systeem op nationale schaal. De effectiviteit van de inzet van het CAS is onderwerp van discussie. In het buitenland is de inzet van PP-systemen niet (meer) vanzelfsprekend. Het Duitse Bundesverfassungsgericht bijvoorbeeld bestempelde begin 2023 een PP-systeem als onconstitutioneel.

De Algemene Rekenkamer stelde in 2022 dat de politie het CAS niet goed op bias en andere risico's controleert. Door bias krijgen sommige mensen niet genoeg bescherming en worden anderen onevenredig gesurveilleerd. Bias kan ontstaan doordat de data die algoritmes gebruiken een vertekend beeld geven. Onbedoelde discriminatie is daarvan een mogelijk gevolg. Het CAS gebruikt geen data uit 'haalincidenten' (actief opgespoorde criminaliteit), om de vertekening door een positieve feedbackloop via het eigen optreden te beperken. Het CAS voorspelt van welke misdrijven aangifte gedaan zal worden (voornamelijk) op basis van eerder gedane aangiftes en openbare data van het Centraal Bureau voor de Statistiek (CBS). Zoals huishoudenssamenstelling, inkomenscijfers en geslachtsverdeling. Dat voorkomt echter niet alle bias. Aangiftebereidheid verschilt bijvoorbeeld per wijk en kan ook weer beïnvloed worden door de inzet van het systeem (feedbackloop).
Het CAS neemt ook mee of er bekende verdachten in de buurt wonen, waardoor 'haalincidenten' eventueel tóch

meetellen. Openbare CBS-data kunnen een 'proxy' (misleidende voorspeller) voor bepaalde etnische groepen vormen. Dit kan discriminatoire politie-inzet tot gevolg hebben.

Meer transparantie is een eerste stap naar betere verantwoording. Proactieve externe transparantie kan helpen om risico's in beeld te krijgen en de effectiviteit of het maatschappelijk nut van de inzet van een PP-systeem inzichtelijk te maken. Een eerste stap is bijvoorbeeld het actualiseren en periodiek bijhouden van een openbaar overzicht van de variabelen die in het CAS gebruikt worden (en hun kalibratie), in het verlengde van de eenmalige openbaarmaking in 2021 op basis van een Wob-verzoek.

Ook van belang is verantwoording over de afweging van de toegevoegde waarde van algoritmische oplossingen. Het is voor organisaties verleidelijk om, op zoek naar een verbetering van processen om bepaalde doelen te bereiken (bijvoorbeeld: minder criminaliteit), ten prooi te vallen aan de gedachte dat technologische innovaties in de kern neutraal zijn. En dat ze, wanneer ze ingezet mogen worden, enkel kansen bieden. Dit wordt ook wel ***technologisch chauvinisme*** genoemd. Door tijdens de besluitvorming nadrukkelijk stil te staan bij de risico's van de inzet van algoritmische systemen, ontstaat een explicietere afweging. Daarover kan of moet de organisatie ook verantwoording afleggen.

Veel handhavingsalgoritmes kennen echter weinig risico's, werken goed en hebben positieve impact. Algoritmegebruik om routinetaken te automatiseren komt vaak voor binnen handhavingsorganisaties en heeft een lange, succesvolle gebruikshistorie. Binnen het handhavingsdomein is het daarom van belang om te focussen op de algoritmes die specifiek risico's voor publieke belangen en grondrechten met zich meebrengen.

Betalingsverkeer

Inzet van algoritmes voor monitoring van het betalingsverkeer brengt de verantwoordelijkheid met zich mee om (groeps)discriminatie te voorkomen. Ook voor financiële dienstverlening worden algoritmes ingezet om efficiënter het betalingsverkeer en mogelijk ongewenste transacties te monitoren. In de financiële sector worden de huidige ontwikkelingen verwelkomd vanwege de wettelijke verplichting om het Nederlands betalingsverkeer op een gedegen manier te monitoren op illegale activiteiten. Zo zijn er risico's op witwassen, financiering van terrorisme en algemenere fraude waarop banken alert moeten zijn en waarnaar zij actief op zoek moeten.

Monitoring van elke individuele transactie is door de omvang van het digitale betalingsverkeer niet mogelijk. Door algoritmes met patroonherkenning toe te passen op basis van data over ongebruikelijke transactiepatronen, modus operandi en risico-indicatoren, kunnen soortgelijke (ongebruikelijke) transacties snel worden herkend. Vermoedelijk illegale transacties kunnen door het systeem worden 'gepauzeerd', wat de transactie bevriest. Een medewerker van de bank bekijkt vervolgens –handmatig – of de inschatting van het algoritme terecht is. Bij een ***false positive***, een onterechte aanmerking als mogelijk illegale transactie, merkt de medewerker de melding van het algoritme als foutief aan. De transactie kan dan alsnog doorgaan. Is de melding van het algoritme mogelijk wel terecht, dan kan er verder worden onderzocht of de transactie inderdaad in strijd is met toepasselijke wetgeving.

Inzet van algoritmes voor monitoring is echter niet zonder risico's. Waar algoritmes ingezet worden om ongebruikelijke patronen te herkennen, bestaat ook snel het risico dat bias kan resulteren in ongewenste discriminerende effecten. Het is essentieel om vooraf mogelijke risico's voor publieke waarden en grondrechten in kaart te brengen. Maar ook om tijdens de inzet bekende en onvoorziene risico's te identificeren en te beheersen. Van belang is dat ook hier, net als bij andere toepassingen van technologie in de samenleving, mogelijk een chilling effect kan optreden. Dit dient ook meegenomen te worden bij het identificeren en beheersen van risico's. Adequate risicobeheersing kan voorkomen dat het inzetten van algoritmes met een hoog risico voor publieke waarden en grondrechten ook daadwerkelijk groepen of individuen treft. De Nederlandsche Bank (DNB) onderzoekt of financiële instellingen beleid en procedures hebben en maatregelen nemen om het risico op discriminatie te beheersen.

Gemeentelijke sociale voorzieningen

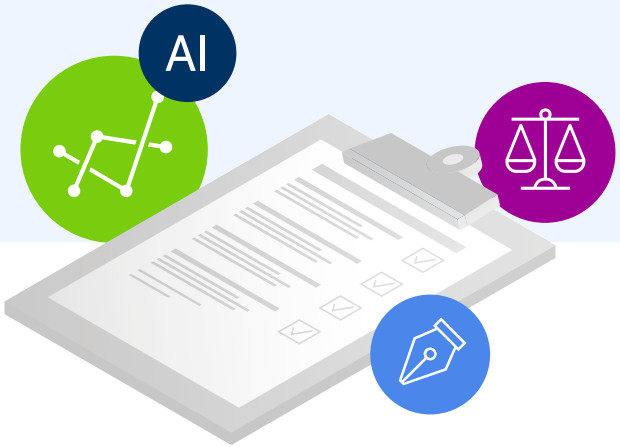
Onvoldoende volwassenheid in sommige gemeentelijke organisaties staat verantwoorde inzet van algoritmes voor belangrijke sociale voorzieningen in de weg. Diverse media hebben bericht over terugkerende problematiek bij algoritmegebruik voor onder meer fraudevoorspelling in de socialevoorzieningensector. Ook na de uitspraak over het Systeem Risico Identificatie (SyRI) van februari 2020 door de Rechtbank Den Haag blijven gemeentes door heel Nederland algoritmes gebruiken voor het opsporen van frauderisico's. Deze systemen en toepassingen variëren in complexiteit, maar kunnen in veel gevallen een grote impact hebben.

Veel van deze systemen gebruiken algoritmes om op basis van historische data de kans op fraude door individuele ontvangers of bepaalde groepen ontvangers in te schatten. Ervaringen uit het verleden, perceptie van risico's, maar ook indicatoren zonder bekende herkomst of waarde worden ingezet om personen of groepen als potentiële fraudeurs aan te merken. Zo kan een beroep als kapper een hoger frauderisico opleveren dan advocaat. Of worden mensen met een koophuis veel minder vaak aangemerkt als mogelijk fraudeur dan mensen die in een huurhuis of een woonwagen wonen.

De inzet van deze systemen die het potentiële frauderisico van een individu inschatten kan impact hebben op levens van individuen, gezinnen en hele groepen in onze samenleving. Aangemerkt worden als mogelijke fraudeur, kan mensen grote emotionele en financiële schade toebrengen. Mensen zijn bij voorbaat verdacht en kunnen door de ondoorzichtigheid van gebruikte systemen maar moeilijk achterhalen waarom ze als fraudeur worden aangemerkt en wat ze daartegen kunnen inbrengen. De schaal waarop dit gebeurt, kan die individuele schade ook vertalen in substantiële maatschappelijke schade. Dat blijkt uit de recente geschiedenis bij toeslagen en SyRI.

De risico's van algoritmes om fraude te voorspellen zijn groot en hun gebruik vereist daarom voldoende *checks and balances*. Gemeentelijke organisaties moeten ingericht zijn om voortdurend te controleren op de risico's van dergelijk algoritmegebruik. De hiervoor benodigde checks and balances moeten zien op zowel de ontwikkeling als de inzet van zulke systemen. Risico's kunnen namelijk in beide fases ontstaan en opgemerkt worden.

In zowel de ontwikkelings- als de gebruiksfase van systemen om fraude te voorspellen moet het nut van een algoritme worden afgewogen tegen de risico's voor grondrechten en publieke waarden. Systemen die fraude voorspellen brengen grote risico's mee voor publieke waarden en grondrechten. Daarom moeten ze voortdurend in overweging genomen worden. Waar nodig moeten passende maatregelen worden getroffen. Soms kan het afzien van de inzet van deze systemen met een hoog risico een legitieme uitkomst zijn van een weloverwogen besluitvormingsproces, als blijkt dat de voordelen onvoldoende opwegen tegen de risico's voor publieke waarden en grondrechten. Ook hier moet dus worden gewaakt voor het eerdergenoemde technologisch chauvinisme.



3. Beleid en regelgeving

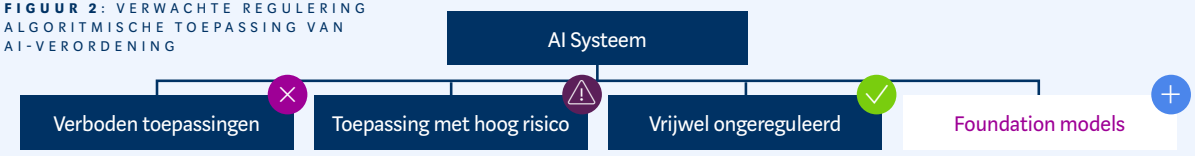
Het opbouwen van nieuwe, aanvullende wet- en regelgeving voor de beheersing van algoritmerisico's is in volle gang en is een mondiale uitdaging.

Internationale AI-kaders, zoals van UNESCO en de OESO, geven richting aan de eisen die gesteld moeten worden. Het is nu zaak dit door te vertalen naar bindende (inter)nationale wetgeving om extra waarborgen te creëren in aanvulling op al bestaande relevante wet- en regelgeving.

Internationaal

De AI-verordening zal algoritmische systemen en toepassingen verbieden of sterker reguleren. Systemen en toepassingen die de vrije keuze van mensen te veel beperken, mensen uitbuiten of mensen manipuleren, worden verboden. Daarnaast reguleert de AI-verordening een aantal 'hoog risico'-systemen. Vrijwel alle specifiek te reguleren toepassingen van hoog risico-systemen die worden genoemd in de AI-verordening, hebben effect op publieke waarden en grondrechten, zoals het recht op gegevensbescherming of gelijke behandeling. Bijvoorbeeld op het gebied van biometrie, werving en selectie en onderwijs. Maar ook systemen die worden gebruikt in kritieke infrastructuur of die zijn toegevoegd aan andere systemen, zoals auto's en medische hulpapparatuur. Dergelijke systemen moeten aan aanvullende voorwaarden voldoen. Het gaat dan bijvoorbeeld om voorschriften voor transparantie, menselijk toezicht en juistheid van data. Van belang is dat ontwikkelaars de risico's van algoritmes voortdurend in de gaten houden en dat zij risico's voor de veiligheid, grondrechten en andere publieke belangen beperken. Ook de organisaties die deze systemen in de praktijk gaan inzetten (gebruikers), zullen zich aan regels moeten houden. Het is mogelijk dat ze verplicht een ***impact assessment*** voor grondrechten moeten gaan uitvoeren. Daarnaast worden in de AI-verordening waarschijnlijk aanvullende regels opgenomen over ***general purpose AI*** en ***foundation models*** (o.a. generatieve AI). Zie figuur 2.

FIGUUR 2: VERWACHTE REGULERING ALGORITMISCHE TOEPASSING VAN AI-VERORDENING



Bepaalde AI-toepassingen worden expliciet verboden
Voorbeeld: AI-toepassingen voor social scoring

Een brede groep van AI-toepassingen die classificeren als hoog risico moeten voldoen aan specifieke regelgeving voorafgaand en tijdens gebruik
Voorbeeld: AI-toepassingen voor het filteren van sollicitaties

AI-systemen die interactie hebben met personen (of content generen) moeten zichzelf als AI-system kenbaar maken

Onderhandelingsvraagstuk is of er een aparte categorie komt met specifieke verplichtingen voor aanbieders van foundation models (e.g. GPT4)
Voorbeeld: specifieke data-governance-eisen, mede ter verificatie van de geschiktheid van de databronnen

Toezicht op de AI-verordening zal waarschijnlijk voor een belangrijk deel nationaal worden ingericht. In Nederland zijn de gesprekken hierover al van start gegaan.

De AP verwelkomt de toenemende aandacht voor het beschermen van grondrechten en fundamentele vrijheden in de onderhandelingen over de AI-verordening. Concreet betekent dit dat ontwikkelaars moeten zorgen dat ze risico's voor bijvoorbeeld gegevensbescherming, discriminatie, democratie en rechtsstaat in de gaten houden en aanpakken. En toezichthouders zullen moeten nagaan of dat goed gebeurt. De onderhandelingen over de AI-verordening zijn volop aan de gang. Een definitief politiek akkoord rond eind 2023, en een snelle implementatie, zou een belangrijk resultaat zijn om de razendsnelle innovatie tegemoet te komen met regelgeving die publieke waarden en grondrechten effectief beschermt.

Tegelijkertijd mogen we van de AI-verordening geen wonderen verwachten, met name omdat bij systemen met een hoog risico de wettelijke compliance primair leunt op de eigen inschatting van producenten. Producenten van hoog-risico-systemen voor bijvoorbeeld werving en selectie, onderwijsbeoordelingen, fraudedetectie maar ook ***predictive policing***, hoeven een systeem niet extern te laten toetsen voordat zij dit op de markt kunnen brengen. Een eigen beoordeling van compliance met essentiële eisen en standaarden volstaat. Hierdoor blijft de kans nadrukkelijk bestaan dat niet-geschikte algoritmische systemen met hoog risico op de markt komen en ingezet worden door private en publieke partijen. Een dergelijk systeem verdwijnt pas van de markt als de toezichthouder constateert dat het AI-systeem niet in orde is. Mogelijk is dan al schade aangericht.

Het toekomstige AI-verdrag zal internationaal de toon zetten voor AI-regulering. Er is ook een AI-verdrag in de maak. Een eerste versie van dit verdrag werd door de Raad van Europa gedeeld in januari 2023. Het verdrag bouwt voort op het Europese verdrag voor de Rechten van de Mens (EVRM) en Conventie 108 over geautomatiseerde verwerking van persoonsgegevens. Aangesloten landen moeten er onder meer voor zorgen dat grondrechten worden beschermd als overheidsorganisaties algoritmes gebruiken in hun besluitvorming. Ook moeten deze landen waarborgen dat individuele vrijheid, menselijke waardigheid en autonomie, democratische processen, volksgezondheid en het milieu worden beschermd als overheden, maar waarschijnlijk ook private partijen, algoritmes gebruiken. Daarnaast bevat het verdrag principes voor bijvoorbeeld gelijkheid, privacy, transparantie en toezicht. Verder schrijft het verdrag maatregelen voor, bijvoorbeeld voor rechtsbescherming en menselijke beoordeling. Omdat het verdrag ook buiten de EU zal gelden en landen als de Verenigde Staten en Japan betrokken zijn, heeft het verdrag de potentie een internationale standaard te worden voor de ontwikkeling en het gebruik van algoritmes in zowel de publieke als de private sector.

Nationaal

Op nationaal niveau is er een duidelijke ambitie om algoritmerisico's beter te beheersen, te beginnen bij de overheid. Zo wordt er gestreefd naar het uitvoeren van mensenrechtentoetsen op algoritmes en wordt gewerkt aan een implementatiekader voor de inzet van algoritmes. De AP beoordeelt beide ambities als goede stappen om duidelijkheid te bieden aan overheidsorganisaties. Het helpt ook dat er één duidelijk toetsingskader is waarnaar de overheid nadrukkelijk verwijst, het Impact Assessment Mensenrechten en Algoritmes (IAMA). Wel wijst de AP op: (i) een proportionele toepassing van dergelijke toetsingsmechanismen en kaders; en (ii) een risicogeoriënteerde infasering door organisaties. Het moet voor organisaties proportioneel en behapbaar zijn. Het doorlopen van het IAMA is een zorgvuldig, veelomvattend maar ook tijdrovend proces. Niet voor elk type algoritme is dit proportioneel. Het is nadrukkelijk wel proportioneel voor de belangrijkste algoritmes die onderdeel zijn van maatschappelijk kritische processen die samenhangen met de publieke taken van overheidsorganisaties. In de optiek van de DCA is de eerste stap dus dat organisaties hun belangrijkste algoritmes rangschikken en daarbij een mensenrechtentoets uitvoeren en dat zij implementatiekaders langslopen. Overheidsorganisaties moeten hiervoor algoritmes met een hoog risico in kaart brengen, wat ook noodzakelijk is in aanloop naar de AI-verordening. Naar verwachting zal de definitieve verordening begin 2024 in werking treden. De AI-verordening zal na een nog vast te stellen periode (2-3 jaar) van toepassing worden. Sommige overheidsalgoritmes zullen onder de AI-verordening als hoog risico worden geclassificeerd. Een volledig overzicht van algoritmes die ingezet en ontwikkeld worden, met bijbehorende risico's en classificatie, zal in de eerste helft van 2024 tot stand moeten komen. Dit is nodig om voorbereidingen te kunnen treffen voor verplichtingen onder de AI-verordening en om deze algoritmes in het algoritmeregister te registreren. De DCA zal nadrukkelijk de voortgang binnen de overheid volgen en in volgende rapportages hierover rapporteren.

Voor startende ondernemingen actief op algoritmeterrein is er publiek-private samenwerking die kan helpen met zorgvuldige productontwikkeling en compliance met regelgeving. Een voorbeeld zijn de ELSA-labs. Dit zijn een soort kraamkamers, waarin er bij de ontwikkeling van algo-

ritmische systemen vanaf het begin expliciet aandacht is voor de ethische, juridische en maatschappelijke belangen. Startende ondernemingen zijn niet belast met (transitie-) uitdagingen bij bestaande systemen. Dergelijke labs bieden ruimte om op een gecontroleerde manier op kleine schaal een systeem te ontwikkelen en vanuit de praktijk lessen te trekken. Bijvoorbeeld door gebruik te maken van synthetische data. Daarmee wordt beheerste innovatie mogelijk. De samenwerking tussen overheidsorganisaties, kennisinstellingen en de private sector helpt daarbij.

De AP is aangewezen als coördinerende partij in het algoritmetoezicht. Om invulling te geven aan deze nieuwe taak is de DCA begin 2023 van start gegaan.
Toezicht in Nederland is thematisch en sectoraal georganiseerd. Sectoraal toezicht brengt veel specifieke kennis met zich mee, maar ook het gevaar dat overkoepelende risico's voor meer algemene publieke waarden en grondrechten onderbelicht blijven. Colleges, markttoezichthouders en rijksinspecties werken steeds vaker en intensiever samen in het toezicht op algoritmes, mede via werkgroepen. Het doel is om kennis te delen en gezamenlijk als externe toezichthouders grip te krijgen op algoritmes en de effecten van de inzet van algoritmes. In de tweede helft van 2023 stelt de DCA via een uitvraag onder toezichthouders een totaalbeeld op van de wijze waarop vanuit alle verschillende toezichtrollen tegen algoritmerisico's wordt aangekeken.

Toezichthouders en marktpartijen bereiden zich voor op nieuwe regelgeving. Een snelle en eenduidige aanpak is nodig om nieuwe regelgeving te laten aansluiten bij bestaande regelgeving, uitleg te bieden en het toezicht nationaal juist in te richten en aan te laten sluiten bij andere lidstaten. Voor systemen en toepassingen binnen de private sector is het van belang om naast bestaande regelgeving en standaarden oog te houden voor publieke waarden en grondrechten. En deze te borgen in aanloop naar en in de geest van nieuwe regelgeving, zoals de AI-verordening. De wijze van invulling kan per toezichtgebied en sector verschillen en vraagt van toezichthoudende organisaties eenzelfde innovatieve aanpak als bij de inzet van nieuwe technologie.

De AP is positief over het algoritmeregister en ziet ruimte voor een verplichte, maar risicogebaseerde versnelling van de vulling van het register. Het algoritmeregister is de basis voor overzicht en inzicht in de algoritmes die overheidsorganisaties gebruiken. Positief is dat het register ook concreet beschrijft hoe en waarom een organisatie het algoritme toepast. Eind juni 2023 waren ongeveer 120 algoritmes geregistreerd, voornamelijk afkomstig vanuit enkele grotere Nederlandse gemeenten. Aandachtspunt is dat het register op dit moment ook wordt gevuld met algoritmes die beperkt risico met zich meebrengen, zoals een systeem om automatisch documentnummers aan besluiten toe te voegen. In eerste instantie zou de focus moeten liggen op systemen met een hoog risico. De classificering kan daarbij voorlopig op basis van selfassessment plaatsvinden, aan de hand van de (voorlopige) lijst van hoog-risico-systemen onder de AI-verordening. De AP is voorstander van verplichte registratie van hoog-risico-systemen in het algoritmeregister. Er moet een deadline komen voor eerste vulling van het register. De ambitie mag hier groot zijn, wederom met de gedachte in het achterhoofd dat het in eerste instantie het belangrijkste is dat de basisgegevens over de risicovolste algoritmes zo snel mogelijk worden geregistreerd.
Dit geeft ook toezichthouders vervolgens een objectief vertrekpunt om het gesprek met organisaties aan te gaan.

Oprichting DCA

Per 2023 is de Autoriteit Persoonsgegevens (AP) gestart als coördinerende autoriteit voor het toezicht op algoritmes. De AP geeft invulling aan deze rol vanuit de nieuwe directie Coördinatie Algoritmes (DCA). De toewijzing van deze taak aan de AP komt voort uit de ambitie om bij de ontwikkeling en inzet van algoritmes publieke waarden en grondrechten beter te beschermen.

De focus van de DCA ligt op het beter beschermen van publieke waarden en grondrechten. Zoals discriminatie en willekeur voorkomen en transparantie bevorderen. Verder kijkt de DCA naar eerlijkheid van algoritmes en het voorkomen van misleiding. De werkzaamheden van de DCA staan los van (de intensivering van) het AP-toezicht op algoritmes die persoonsgegevens verwerken.

Risicosignalering is een belangrijk onderdeel van de activiteiten van de DCA. Deze activiteit hangt samen met het versterken van de samenwerking in het toezicht en het bevorderen en faciliteren van gezamenlijke normuitleg en ***guidance*** voor organisaties. De focus bij risicosignalering ligt op het signaleren en analyseren van sectoroverstijgende en overkoepelende risico's en effecten van algoritmes. En op ontwikkelingen in beleid en regelgeving.

De DCA werkt aan netwerken en structuren om signalen te ontvangen en te bespreken. In haar coördinerende rol beoogt de DCA in verbinding te staan met toezichthouders, sectorvertegenwoordigers, uitvoeringsorganisaties, belangenorganisaties, het maatschappelijk middenveld, de wetenschap en specialistische organisaties en instituten. De DCA bouwt deze netwerken op en richt rapportagestructuren in. Ook internationale netwerken zijn een belangrijke bron, om risico's die internationaal worden opgemerkt onder de aandacht brengen. Door de signalen door de tijd heen te blijven monitoren kan de DCA trends identificeren. De RAN biedt een periodiek overzicht van deze risico's en effecten, zowel in het overkoepelende deel als in het deel waar met concrete voorbeelden algoritmerisico's worden besproken die zijn gesignaleerd op basis van media, stakeholders, internationale monitoring of vanuit de toezichtpraktijk. In de RAN bespreekt de AP deze concrete voorbeelden nadrukkelijk niet vanuit haar taak als AVG-toezichthouder.

Samenwerking met toezichthouders is essentieel. Zowel nationaal als internationaal worden bestaande samenwerkingen geïntensiveerd en waar nodig aanvullende samenwerkingen opgezet. Het landschap van toezicht in Nederland is omvangrijk, en wordt gekenmerkt door de sterke sectorale en thematische kennis. Dit kan door samenwerking optimaal worden benut. Om samen voor effectief toezicht te zorgen zal de DCA het landschap van toezicht op algoritmes in kaart gaan brengen. Hierdoor moet inzichtelijk worden welke toezichthouders een rol spelen, welke taken er nodig zijn, waar deze belegd zijn, maar hierbij wordt ook gekeken naar mogelijke hiaten in toezicht.

De DCA werkt ook aan een visie op algoritmerisico's. Om de beoordeling van risico's te ondersteunen werkt de DCA aan een kader waarmee gekeken kan worden naar het type algoritmesystemen, beheersingssystemen en de wijze waarop algoritmes ontwikkeld en ingezet worden. De combinatie van deze drie elementen is bepalend voor de mate waarin risico's voor waarden en grondrechten als discriminatie, eerlijkheid, uitlegbaarheid en misleiding kunnen materialiseren. De DCA verwacht in de tweede helft van dit jaar een eerste visiedocument ter consultatie op te leveren.

AUTORITEIT
PERSOONSGEGEVENS