AI-TOEPASSINGEN VOOR VEILIGHEID, VREDE EN RECHT

Effectief, veilig en mensgericht

NL AI Coalitie

INHOUDSOPGAVE

1. INTRODUCTIE

Artificiële Intelligentie (AI) valt niet meer weg te denken uit onze samenleving. Het wordt volop toegepast. In een digitaliserend tijdperk met toenemende dreigingen biedt de inzet van AI-kansen als het gaat om het versterken van onze fysieke en digitale veiligheid. AI kan bijdragen aan het beschermen van de burger, het waarborgen van het recht en het vergroten van het algemeen welzijn. Tegelijkertijd brengt de inzet van AI ook risico's met zich mee en dient er balans gevonden te worden tussen de effectieve inzet van AI-toepassingen en het beschermen van de rechten van burgers.

De werkgroep Veiligheid, Vrede en Recht (VVR) van de Nederlandse AI Coalitie brengt relevante nationale spelers bij elkaar om effectieve, veilige en verantwoorde AI-toepassingen te ontwikkelen en in te zetten voor veiligheid, vrede en een goedwerkend rechtssysteem waarbij de mens centraal staat. Zo hoopt zij Nederland tot koploper op het gebied van mensgerichte AI voor veiligheid, vrede en recht te maken. In deze publicatie laat de werkgroep zien wat daarvoor nodig is, waar haar focus ligt en waar zij momenteel staat. Daarbij worden verschillende oplossingen, initiatieven en toepassingen uitgelicht.

1.1 WERKGROEP VEILIGHEID, VREDE EN RECHT

De werkgroep richt zich op de inzet van AI binnen het brede veiligheidsdomein. Hierbij kun je bijvoorbeeld denken aan het anticiperen op criminaliteit, forensisch onderzoek, fysieke veiligheid, crisisbeheersing, de rechtspraak en openbaar bestuur. Zij streeft een mensgerichte aanpak na waarbij AI vanaf het moment van ontwikkeling op een veilige en verantwoorde manier wordt ingezet. De focus van de werkgroep ligt op privacy-bestendig data delen, toepassingen ten behoeve van beslisondersteuning, cybersecurity en taal- en spraaktechnologie. Uitgangspunt is een mensgerichte aanpak waarbij uitlegbaarheid, transparantie en inclusie centraal staan. De werkgroep zet zich in om Nederland zowel op strategisch als operationeel-tactisch niveau leidend te laten zijn in artificiële intelligentie voor veiligheid, vrede en recht. De werkgroep heeft hierbij aansluiting gezocht bij de AI-strategie van het ministerie van Justitie en Veiligheid.

1.2 TRIPLE HELIX

De werkgroep bestaat uit 110 deelnemers die de triple helix vertegenwoordigen uit iedere regio in Nederland: (semi)overheid (zowel beleids- als uitvoeringsorganisaties), kennisinstituten (academisch en toegepast onderzoek, onderwijsinstellingen), bedrijfsleven (variërend van start ups, scale ups tot grote bedrijven) en maatschappelijke organisatie en ngo's. Lopende en nieuwe AI-projecten worden besproken, evenals bestaande praktijkvoorbeelden waarin AI verantwoord wordt ingezet (use cases). Projectplannen worden opgezet en besproken, en de projecten worden met elkaar in verbinding gebracht. Rondom de projecten worden consortia gevormd van stakeholders die het project dragen en verder tot ontwikkeling brengen met behulp van financieringsbronnen zoals het Nationaal Groeifonds en het Missiegedreven Innovatieprogramma 'Data en intelligence'. Barrières die een verantwoorde inzet van AI belemmeren, worden door alle triple helix partijen binnen de werkgroep gezamenlijk geslecht.

De Nederlandse AI Coalitie is hét samenwerkingsverband van betrokken stakeholders op het gebied van AI in Nederland. Zij kent een publiek-private samenstelling waarbij overheid, bedrijfsleven, onderwijs- en onderzoeksinstellingen en maatschappelijke organisaties zich inzetten om AI-ontwikkelingen te versnellen en AI-initiatieven met elkaar te verbinden. De NL AIC bestaat uit werkgroepen die zich richten op verschillende bouwstenen en focusgebieden die cruciaal zijn voor de ontwikkeling en toepassing van AI in Nederland en daarbuiten.

1.3 VERSTERKING INITIATIEVEN

De strategische lijnen van de werkgroep VVR vallen samen met de Kennis- en innovatieagenda Veiligheid 2020-2023 (KIA). Deze agenda komt voort uit het topsectorenbeleid van de overheid, waarbij wordt ingezet op missiegedreven innovatiebeleid voor vier maatschappelijke thema's. Veiligheid is een van deze maatschappelijke thema's. Binnen dit thema zijn verschillende missies geformuleerd, zoals de missie 'data en intelligence':

"in 2030 verzamelen veiligheidsorganisaties nieuwe en betere data, met slimmere analyses worden de juiste interventies gedaan en worden ze niet verrast." Hiervoor is een meerjarige missiegedreven innovatieprogramma (MMIP) opgesteld, met aandacht voor privacy-bestendige informatiedeling en beslissingsondersteuning. Door een koppeling te maken met de KIA wordt dubbel werk voorkomen en versterken beide initiatieven elkaar.

1.4 VERBINDENDE FACTOR

De werkgroep bouwt een ecosysteem door alle relevante stakeholders bijeen te brengen. Zij legt verbindingen tussen alle relevante stakeholders op het gebied van veiligheid en AI, en ondersteunt de veiligheidssector in de volle breedte met een verantwoorde ontwikkeling en implementatie van AI-gerelateerde technologieën. Het idee voor een eigen werkgroep rond dit thema is ontstaan vanuit het Data Science Initiative dat is opgezet door de gemeente Den Haag om de waarde van data science en artificiële intelligentie te benutten voor veiligheid, vrede en recht. Dit initiatief is inmiddels opgegaan in de werkgroep VVR en verder verankerd in de AI-hub Zuid-Holland.

De werkgroep zorgt ervoor dat alle relevante spelers met elkaar afstemmen en samenwerken op inhoud en resultaat. In opdracht van het ministerie van Justitie en Veiligheid heeft Security Delta (HSD) in zowel de werkgroep VVR als de MMIP 'Data en Intelligence' de rol van voorzitter en coördinator op zich genomen. HSD stimuleert en faciliteert kennisdeling en verbindt partijen volgens het HSD-samenwerkingsmodel: vraaggestuurd, doelgericht en op basis van vertrouwen. Door het HSD-netwerk te koppelen aan de werkgroep VVR en de KIA Veiligheid kunnen initiatieven worden verbonden, versterkt en opgeschaald.

Financiering

Vanuit de inhoudelijke ambities van projecten en consortia wordt gezocht naar passende (co-) financiering. Dit kan gaan om calls vanuit het Nationaal Groeifonds en NWO, maar ook om de inzet van RVO-instrumenten en regionale, Europese of andere internationale programmafinanciering.

Om de Nederlandse positie te versterken en de kansen te verzilveren, is het meerjarige AiNed-programma opgesteld door de Nederlandse AI Coalitie. Het programma versnelt de ontwikkeling en toepassing van AI, zodat Nederland economisch en maatschappelijk de vruchten van AI kan plukken en internationaal met de koplopers mee kan doen. Het meerjarige AiNed-programma vraagt een investering van 1,05 miljard euro uit het Nationaal Groeifonds voor de periode 2021-2027, terwijl eenzelfde bedrag wordt geïnvesteerd door private partijen (bedrijven) en overheden. De totale omvang van het programma is 2,1 miljard euro. De Nederlandse overheid erkent het belang van AI en de kracht van de NL AIC en haar deelnemers, en heeft in april 2021 een bedrag van 276 miljoen euro vanuit het Nationaal Groeifonds aan de eerste fase van het meerjarige AiNed-programma toegekend.

De werkgroep VVR sluit aan bij de Kennis- en Innovatie-agenda Veiligheid (KIA). Via de NWO KIC call 'Data en Intelligence' draagt zij bij aan de overkoepelende doelstelling van deze kennis- en Innovatieagenda en de onderliggende meerjarige Missiegedreven Innovatie Programma's. Deze call biedt de mogelijkheid voor inter-disciplinaire consortia van kennisinstellingen, publieke en private partners om voorstellen in te dienen met een totale omvang van minimaal €750.000 en maximaal €3.000.000 per project of consortium. De call heeft als doel om alfa-, bèta- en gammawetenschappers samen een bijdrage te laten leveren aan betere en bruikbare intelligenceproducten die voorzien in de behoeften van intelligence- en veiligheidsprofessionals voor operatio-nele, tactische en strategische taken.

1.5 NORMEN EN WAARDEN

De ontwikkeling van AI voor veiligheid, vrede en recht is een samenspel tussen de inzet van effectieve technologie en de normen en waarden die horen bij een democratische digitale rechtsstaat. Bij een verantwoord gebruik van AI staat niet de techniek centraal, maar de vraag. In het geval van de werkgroep VVR gaat het dan om de vraag en behoefte van veiligheidsprofessionals met aandacht voor de belangen van burgers. De werkgroep wil een bijdrage leveren aan betere en bruikbare inlichtingenproducten (intelligence) op basis van artificiële intelligentie, die aanvullend en van meerwaarde zijn bij operationele, tactische en strategische taken. In Nederland vinden we bepaalde normen en waarden belangrijk. Deze dienen dan ook verdisconteerd te worden in het gebruik van AI. Dit is alleen mogelijk als alle betrokken stakeholders met elkaar de normen, waarden en belangen in hun context beoordelen. De normenkaders die hierbij relevant zijn, zijn gebaseerd op een mensgerichte aanpak van AI (technologie dient ten dienste te staan aan de mens), zoals privacy, maatschappelijke acceptatie, soevereiniteit, inclusie, transparantie en uitlegbaarheid.

1.6 BETROUWBAAR EN CONTROLEERBAAR

De kracht van AI is dat zij met behulp van slimme algoritmes functies kan uitvoeren die normaliter worden geassocieerd met het menselijk brein, zoals het oplossen van problemen en het herkennen van patronen. AI is in staat menselijke denkkracht te automatiseren. Zij werkt zeer efficiënt, snel en 'ziet' dingen die aan mensen voorbijgaan. In bijvoorbeeld forensisch onderzoek en recherchewerk is AI dan ook een welkome aanvulling. Het gebruik van deze technologie is echter niet zonder risico. Zo maakt AI gebruik van data, en data kan gemanipuleerd worden. Data kan ook 'vervuild' zijn met verkeerde of onvolledige gegevens, waardoor AI verkeerde dingen leert en dus verkeerde uitkomsten genereert. Dit kan ook het gevolg zijn van aannames en vooroordelen van de programmeurs, die onbedoeld in het AI-systeem worden ingebouwd. Om deze reden is het van groot belang om op transparante wijze AI toe te passen en inzicht te geven in de gebruikte data en algoritmes. De algoritmes en de concrete uitkomsten moeten daarom betrouwbaar en controleerbaar zijn. Hiermee belanden we technologisch gezien aan de kant van explainable AI en tegelijkertijd op het terrein van de ethiek. De werkgroep VVR maakt zich daarom sterk voor een adequate en transparante inzet van AI, ter bevordering van de welvaart en het welzijn in Nederland.

1.7 BOUWSTENEN NL AIC

De Nederlandse AI Coalitie (NL AIC) heeft werkgroepen in zogenaamde bouwstenen die randvoorwaardelijk en ondersteunend zijn aan de focusgebieden van de werkgroep. De bouwstenen dragen bij en zijn soms zelfs essentieel om projecten tot een succes maken. Deze werkgroepen per bouwsteen zorgen voor de opbouw van gemeenschappelijke kennis, expertise en oplossingen voor de geconstateerde uitdagingen. Deze uitdagingen spelen in de gebieden waarop de werkgroep VVR zich focust, maar ook in de gebieden van andere sectorale werkgroepen zoals Cultuur en Media, Defensie, Energie en Duurzaamheid, Gezondheid en Zorg, en Technische Industrie. Deze bouwstenen worden hieronder kort beschreven en zullen in hoofdstuk 2 verder worden toegelicht voor het VVR-domein. Omdat data delen gedeeltelijk gaat over het ontwikkelen van privacy-bestendige technologie is er binnen dit position paper voor gekozen om hier een focusgebied van te maken.

1 Human Capital

AI brengt de nodige verandering teweeg op de arbeidsmarkt. Het vervangt taken en maakt werk plezieriger en efficiënter. Er is daarom grote behoefte aan talent dat AI kan programmeren en er in de praktijk mee kan omgaan. Dat vergt opleiding en scholing. Iets waar nu een groot gebrek aan is. Voor Nederland is dit een kans. Als overheid, bedrijfsleven, onderwijs en vertegenwoordigers van de maatschappij goed samenwerken in de planvorming en uitvoering, dan kunnen zij hoogwaardige en vernieuwende opleidings- en scholingstrajecten ontwikkelen die (internationaal) aandacht en mensen trekken. Het is de bedoeling dat een deel van dit buitenlands talent in Nederland aan de slag gaat en dat Nederlandse werknemers goed worden voorbereid op een toekomst met AI in hun werk.

2 Mensgerichte AI

Onze samenleving staat voor een grote verandering. De toepassing van artificiële intelligentie wordt wel een 'systeemverandering' genoemd, omdat het de samenleving ingrijpend zal veranderen, net zoals de industriële revolutie, de uitvinding van computers en het internet hebben gedaan. AI zal de omgang met digitale systemen en diensten enorm sterk beïnvloeden. Hoe moet onze samenleving zich verhouden tot deze digitale transformatie? Hoe borgen we publieke waarden, grondrechten en democratische vrijheden? En hoe zorgen we ervoor dat iedereen kan profiteren van deze voortuitgang? AI dient de mens. Daar is deze bouwsteen op gericht.

3 Data Delen

Artificiële Intelligentie is afhankelijk van data. Hoe meer relevante data er beschikbaar is en hoe hoger de kwaliteit, hoe beter het AI-systeem kan leren. De toegang tot data is dus cruciaal voor de inzet van AI. In Nederland wordt data vaak afgesloten bewaard; niet toegankelijk voor anderen. Meestal gebeurt dit vanwege juridische of commerciële redenen. Om die barrières te doorbreken moet op een goede en verantwoorde manier, veel sneller en beter dan we nu gewend zijn, het delen van data worden georganiseerd. 'Privacy enhancing technologies' zijn hulpmiddelen om datadelen mogelijk te maken. Dit zorgt ervoor dat AI-systemen kunnen leren van data, waardoor een betere, nauwkeuriger en zorgvuldiger dienstverlening mogelijk wordt. Waarden als vertrouwen, inzicht, kennis, privacy- en databescherming, en democratische principes zijn hierbij het uitgangspunt.

4 Research en Innovatie

Een gecombineerde aanpak van fundamenteel en toegepast onderzoek met innovaties in de gehele waardeketen zal AI-onderzoek, -ontwikkelingen en -toepassingen versterken en versnellen. Hiervoor wordt gebouwd aan een AI-netwerk van partners dat voortborduurt op bestaande kwaliteiten en ruimte biedt aan nieuwe initiatieven. Zo is het Innovation Center for Artificial Intelligence (ICAI) al vanaf de oprichting van NL AIC betrokken als toonaangevend nationaal netwerk van wetenschappelijke AI-samenwerkingsverbanden.

5 Startups en Scale-ups

De werkgroep VVR is zich ervan bewust dat haar deelnemers niet de wijsheid in pacht hebben en dat een deel van de oplossingen te vinden is bij startups en scale ups. Zij kijkt dan ook breder naar bedrijven, startups en scale ups uit andere domeinen die van betekenis kunnen zijn voor het verantwoord en mensgericht inzetten van AI. Deze partijen krijgen het podium bij bijeenkomsten van de werkgroep. Zij krijgen de mogelijkheid zich te presenteren, te sparren over vraagstukken en contact te leggen met deelnemers van de coalitie.

Definitie AI

De laatste jaren is AI een veel besproken en onderzocht onderwerp. Het brede toepassingsveld dat AI biedt, en de kansen en risico's die hierbij komen kijken, zorgen ervoor dat het voor veel verschillende partijen een zeer interessant onderwerp is. In de praktijk worden diverse definities en omschrijvingen van AI gebruikt. De definitie van AI die is ontwikkeld door de Europese Commissie is uitgangspunt voor dit position paper. Deze definitie is breed, sterk onderbouwd en biedt een mooie basis om dieper op specifieke eigenschappen van AI in te gaan.

De Europese definitie van AI luidt: "Kunstmatige intelligentie (AI) verwijst naar door mensen ontworpen systemen die, gegeven een complex doel, in de fysieke of digitale wereld optreden door hun omgeving waar te nemen, de verzamelde gestructureerde of ongestructureerde gegevens te interpreteren, te redeneren op basis van de uit deze gegevens afgeleide kennis en te besluiten welke actie(s) het best kunnen worden ondernomen (volgens vooraf bepaalde parameters) om het gegeven doel te bereiken. AI-systemen kunnen ook worden ontworpen om te leren hun gedrag aan te passen door te analyseren hoe de omgeving door hun eerdere acties wordt beïnvloed."

1.8 FOCUSGEBIEDEN

De ambitie van de werkgroep VVR is om Nederland koploper te laten zijn als het gaat om een verantwoorde inzet van artificiële intelligentie in het veiligheidsdomein. Door het opstellen van een langlopend programma en een meerjarige strategie ontstaat de mogelijkheid om prioriteiten aan te brengen in onderzoek en innovaties, over een langere periode te investeren in innovaties en met elkaar tot procedures en afstemming te komen. De werkgroep richt zich op vier focusgebieden:

1 ***Realiseren van oplossingen voor privacy-bestendige informatiedeling***

Er bestaat enorm veel data over mensen, afkomstig van uiteenlopende bronnen, die momenteel niet op voorhand in samenhang geanalyseerd worden als gevolg van onder andere privacybezwaren. Maar daarmee worden ook maatschappelijk gewenste toepassingen belemmerd. De werkgroep zoekt naar technologische oplossingen om data in samenhang te kunnen analyseren met behoud van geheimhouding van die data.

2 ***Oplossingen om op basis van data en intelligence de juiste beslissingen te nemen***

Om de veiligheidsprofessionals en -bestuurders te allen tijde optimaal voor te bereiden op de beslissingen die zij tijdens een interventie moeten nemen, is het zaak zo accuraat mogelijk inzicht te geven in de situatie waarin zij operationeel, tactisch of strategisch moeten handelen. Het kan dan bijvoorbeeld gaan om (nood)situaties waarin een first responder optreedt, maar ook om het analyseren van documenten ten behoeve van een rechtszaak op basis waarvan een rechter uitspraak doet.

3 ***Oplossingen ter verbetering van cybersecurity-technologieën én cyberveilige AI-systemen***

Digitale dreigingen en de gebruikte aanvalsmethoden ontwikkelen zich snel. De werkgroep richt zich met dit focusgebied op het verbeteren van defensieve cybertechnologie. De focus ligt daarbij op het enerzijds ontwerpen van veilige, geautomatiseerde en privacy-vriendelijke systemen en producten, en anderzijds op het ontwikkelen van technieken om systemen en producten weerbaar te maken en te houden.

4 ***De inzet van taal en spraaktechnologie voor veiligheid, vrede en recht***

De meeste algoritmes om gesproken taal om te zetten in tekst worden ontwikkeld voor de grote wereldtalen en zijn in handen van grote tech-bedrijven. Hoewel er ondersteuning bestaat voor het Standaard Nederlands, bestrijken allerlei varianten, dialecten, accenten en atypische spraak een te kleine afzetmarkt om voor hen interessant te zijn. De technologieën die er zijn in het Nederlandstalig taalgebied, zijn ontwikkeld door specifieke partijen die voorzien in hun eigen behoefte. Daardoor kennen deze systemen beperkingen en zijn ze niet goed inzetbaar in situaties waar geen Standaard Nederlands wordt gebruikt. Om niet afhankelijk te hoeven zijn van de buitenlandse big tech-bedrijven, is de behoefte ontstaan voor soevereine en diverse Nederlandse taal-en spraakalgoritmes die toegankelijk zijn voor iedereen.

2. BOUWSTENEN VOOR VEILIGHEID, VREDE EN RECHT

Het boeken van resultaat met AI wordt versterkt als belangrijke randvoorwaarden goed geregeld zijn. Het ontwikkelen van AI-opleidingen en -scholing, het aantrekken van nationaal en internationaal talent, transparantie over de werking van AI, ethische en juridische kaders waarbinnen AI-systemen werken, internationale samenwerking en het privacy-bestendig delen van informatie zijn essentieel om AI op een juiste en verantwoorde manier te kunnen inzetten.

2.1 BOUWSTEEN: HUMAN CAPITAL

Het ontwikkelen en verantwoord toepassen van AI vraagt om een goede basis van onderwijs, scholing en talentontwikkeling. Deze basis is slechts beperkt aanwezig. De werkgroep VVR wil een bijdrage leveren aan het oplossen van het tekort aan opleidingen en nieuw talent in Nederland. Zij wil onderwijsinstellingen stimuleren prioriteit te geven aan het opnemen van AI in de curricula. Ook draagt zij bij aan de landelijke human capital agenda AI, scholing en een leven lang leren.

De ontwikkeling van onderwijs richt zich als eerste op diverse minors, met thema's als 'AI in society', 'AI in research' en 'AI in engineering'. Dit zijn programma's van een half jaar waaraan 1.500 studenten mee kunnen doen. Ook wordt gekeken naar het inbedden van AI in bestaande vakken en het ontwikkelen van nieuwe vakken bij bestaande opleidingen als cybersecurity, forensics en rechten. AI kent een brede toepassing en is dan ook relevant voor veel sectoren en bijbehorende opleidingen. Daarnaast ontwikkelen de onderwijsinstellingen een specifieke AI-opleiding.

Op dit moment zijn er weinig mensen die werken aan de ontwikkeling van AI of er verantwoord mee om kunnen gaan. Als er al talent is, dan worden ze vaak weggekaapt door andere landen of grote bedrijven. Aangezien AI relevant is voor veel sectoren in onze maatschappij, is er een grote vraag naar talent. Daarom wordt er gewerkt aan het terugdringen van het tekort door het ontwikkelen van nieuw onderwijs op het gebied van AI, waaronder minoren, masters en nieuwe curricula. Tevens worden meerdere Internationaal Talent Programma's gestart die tot doel hebben internationaal talent te werven (wonen en werken), werkgevers daarbij te ondersteunen en talent vast te houden.

Een andere oplossingsrichting is het bijbrengen van AI-vaardigheden bij meer mensen, door middel van 'leven lang leren' (trainingen, online onderwijs, AI-tools voor persoonlijke ontwikkeling) of de inzet voor het belang van AI en digitalisering in het algemeen op middelbare scholen. Zo is inmiddels 'De Nationale AI Cursus' beschikbaar op het speciaal ontworpen platform ai-cursus.nl. Deze cursus is gratis, voor iedereen toegankelijk en heeft als doel zoveel mogelijk Nederlanders goed voor te bereiden op een toekomst met AI. Er zijn verschillende cursussen beschikbaar, waaronder een basiscursus, een kindercursus en cursussen voor verschillende beroepssectoren, zoals gezondheid en zorg, en agri en food. Daarnaast zijn er cursussen voor professionals, te volgen via de academie van het Innovation Center for Artificial Intelligence (icai.ai/academy). Ook wordt er gewerkt aan het verbeteren van de mobiliteit van studenten. Curricula worden op elkaar aangesloten, zodat studenten meerdere onderwijstrajecten kunnen volgen, bijvoorbeeld van mbo naar hbo naar wo.

Vacatures, opleidingen en trainingen op Security Talent en Trainingsplatform NL AIC

Op Security Talent worden vraag en aanbod op het gebied van veiligheid bij elkaar gebracht. In de afgelopen jaren zijn duizenden vacatures en stageplaatsen gepubliceerd en hebben meer dan tweehonderd werkgevers van de site gebruik gemaakt. Er bestaat momenteel een brede vraag naar AI-experts binnen onderzoek, overheid en het bedrijfsleven. Ook worden er meer dan vijfhonderd opleidingen en trainingen ontsloten, zoals meerdere bachelor- en masterprogramma's Artifical Intelligence. Voorbeelden hiervan zijn de Nationale AI cursus en Elements of AI. Het Trainingsplatform van de NL AIC is ontwikkeld door de NL AIC werkgroep Human Capital. Deze pagina richt zich op her-, bij- en omscholing in alle sectoren en biedt het AI trainingsaanbod voor en door deelnemers aan. Zo ook de AI for Business & Government-certificering. Dit is een standaard voor professionals op hbo+-niveau die in hun werkomgeving AI willen inzetten. Professionals die de certificering 'AI for Business' hebben behaald, hebben brede kennis van hoe je AI kunt toepassen in de organisatie en hoe je de organisatie kunt inrichten zodat AI kan worden toegepast.

Professionals, werkgevers en organisaties in het veiligheidsdomein zullen zich met opleiding en training moeten voorbereiden op een tijd waarin veiligheidsprofessionals worden ondersteund door krachtige digitale technologie. Denk hierbij aan faciliterende en zelflerende AI-systemen zoals speech-to-text voor de Nederlandse taal, chatbots en virtual agents die emoties kunnen herkennen. Met het Missiegedreven Meerjarige Innovatie Programma (MMIP) 'De veiligheidsprofessional' worden innovatieve en technologiegedreven vormen van opleiden en trainen ontwikkeld, evenals een *21st century skill-set* voor veiligheidsprofessionals.

2.2 BOUWSTEEN: MENSGERICHTE AI

Inwoners van Nederland moeten de juiste keuzes kunnen maken bij het gebruik van AI in hun dagelijks leven en zo mogelijk betrokken worden bij de ontwikkeling van nieuwe AI-diensten. Daarom wordt naar manieren gezocht om samen te leren en te ontdekken wat de beste en meest wenselijke AI-oplossingen zijn. De werkgroep werkt volgens een mensgerichte aanpak van AI, waarbij AI-systemen ethisch verantwoord en maatschappelijk zinvol worden ontwikkeld. VVR is bij uitstek een domein waarbij AI-toepassingen grote gevolgen kunnen hebben voor de burger. Het is dan ook essentieel om de burger bij de ontwikkeling van systemen centraal te stellen en vanaf het begin mee te nemen in het ontwikkelproces. In de werkgroep worden relevante normenkaders besproken, wat deze voor de werkgroepleden betekenen en hoe deze partijen hun systemen hier het beste op in kunnen richten.

Daarnaast stimuleert de werkgroep de mensgerichte aanpak binnen initiatieven en helpt ze partijen om deze principes in de praktijk te brengen. Zo zijn er onder andere meerdere Ethical, Legal and Societal Aspects (ELSA) labs in ontwikkeling voor VVR. Daarmee kan een bijdrage worden geleverd aan de bestaande en opkomende vraagstukken zoals meaningful human control. In het volgende hoofdstuk worden deze verder toegelicht.

2.2.1 EU HUMAN CENTRIC APPROACH

De Europese Unie (EU) maakt zich sterk voor en wil zich onderscheiden met verantwoorde mensgerichte AI. Door de inzet van mensgerichte AI verwacht zij het leven van de Europese burger aanzienlijk te kunnen verbeteren. Daarnaast zal het grote voordelen opleveren voor de samenleving en de economie. Om deze aanpak te versterken en de risico's van AI-producten te beperken heeft de Europese Commissie een verordening opgesteld. Bij het niet navolgen van de regels kunnen boetes worden opgelegd die oplopen tot zes procent van de jaaromzet van een organisatie. Door nu al bij AI-experimenten rekenschap te geven van de komende regelgeving zorgen we voor uitvoerbare systemen en zijn we beter voorbereid als deze van kracht wordt. Vooral voorkomen we zo al problemen, die anders mogelijk te laat aan het licht zouden komen.

De regulering richt zich op drie soorten AI: verboden, high-risk en andere AI-systemen. AI-systemen die personen onbewust op schadelijke wijze beïnvloeden zijn verboden evenals AI-systemen die kwetsbaarheden van personen uitbuiten. Ook AI-systemen die worden gebruikt voor social scoring (het inschatten en beoordelen op betrouwbaarheid en wenselijk gedrag via 'social credit systems') en real-time biometrische systemen voor opsporing en handhaving zijn verboden. Het gros van de regulering heeft betrekking op high-risk systemen. Dit zijn onder andere AI-systemen die gebruikt worden in speelgoed, de luchtvaart, het onderwijs en medische apparatuur, maar ook voor toepassingen in de rechtspraak en opsporing. De meest prominente eis voor high-risk systemen is de verplichting om een conformiteitsbeoordeling uit te voeren voordat het product in de handel wordt gebracht. Voor andere AI-systemen geldt dat mensen geïnformeerd moeten worden over het gebruik van AI en hoe het AI-systeem wordt ingezet.

2.2.2 TOOLBOX ETHISCH VERANTWOORD INNOVEREN VOOR OVERHEDEN

Naast de EU-wetgeving werkt de overheid ook via andere mensgerichte AI-instrumenten om publieke waarden en mensenrechten bij de inzet van AI te waarborgen. Deze instrumenten zullen samengebracht worden in de online

toolbox 'Ethisch verantwoord innoveren voor overheden'. Het betreft de volgende drie instrumenten: 'de Handreiking AI-systeemprincipes voor non-discriminatie'. Dit hulpmiddel is in januari 2021 afgerond en biedt een praktisch toepasbaar ontwerpkader dat ontwikkelaars helpt om al in de ontwikkelfase van een AI-systeem discriminerende patronen in gegevens zoveel mogelijk te identificeren, te voorkomen en te bestrijden. Daarnaast is deze zomer het Impact Assessment Mensenrechten en Algoritmes (IAMA) gepubliceerd dat gebruikt kan worden bij de keuze om wel of niet een AI-toepassing te ontwikkelen en helpt om de ontwikkeling en implementatie vervolgens op een verantwoorde wijze uit te voeren. Als derde instrument is de Code Goed Digitaal Openbaar Bestuur ontwikkeld. De Code geeft aandacht aan de gevolgen van digitalisering voor openbaar bestuur op basis van principes en waarden voor democratie, rechtsstaat en bestuurskracht. In de bijlage is een overzicht te vinden van de relevante kaders voor het gebruik van AI.

2.2.3 PUBLIEKE CONTROLE OP ALGORITMES

De vier grootste steden van Nederland, provincies, politie en Rijkswaterstaat hebben de handen ineengeslagen om in het project 'Publieke controle op algoritmes' gezamenlijke beleidsinstrumenten te ontwikkelen. Algoritmes worden steeds meer ingezet om menselijke taken te ondersteunen of zelfs over te nemen, ook binnen de overheid. Daarom zijn instrumenten nodig die ervoor zorgen dat algoritmes gecontroleerd en burgers beschermd worden tegen bijvoorbeeld verkeerde uitkomsten. Dit laatste speelt ook een rol bij de aanpassing van de richtlijnen uit 2019 over het beschermen tegen risico's van data-analyses. Het belangrijkste doel van de aangepaste richtlijnen is het creëren van transparantie en voorwaarden om mogelijke risico's van de inzet van algoritmen tegen te gaan, zoals uitlegbaarheid (redenen om AI in te zetten) en auditeerbaarheid (controlemechanismen). Deze aanpassingen sluiten aan bij een vragenlijstonderzoek uit september 2021 waaruit blijkt dat driekwart van de ruim duizend ondervraagde Nederlanders thema's als privacy, menselijke controle op het algoritme en de reden waarom het algoritme wordt gebruikt belangrijk vinden. Ze zouden daar graag meer informatie over willen krijgen.

2.2.4 HET AI VREDESPALEIS

De werkgroep werkt onder de naam 'AI Vredespaleis' aan de realisatie van een internationaal forum voor verantwoordelijke AI op het gebied van veiligheid, vrede en recht. De missie van het forum is om bij te dragen aan de ontwikkeling van Europa tot eerste intelligente, rechtvaardige en veilige AI-samenleving in de wereld. Hiervoor wordt een innovatief en internationaal ecosysteem gecreëerd van overheden, bedrijven, kennisinstellingen en maatschappelijke organisaties. Voortbordurend op de missie van het Data Science Initiative, steunt het forum innovatieve projecten op het gebied van data en AI, gericht op veiligheid, vrede en recht. Parallel hieraan brengt het 'AI Vredespaleis' partijen bijeen om de governance vorm te geven en juridisch te borgen. Ook wil het forum het maatschappelijk debat over AI stimuleren.

The Hague Conference on Responsible AI for Peace, Justice and Security.

Deze conferentie in het Vredespaleis in Den Haag draagt bij aan de ontwikkeling van Europa tot eerste intelligente, rechtvaardige en veilige AI-samenleving in de wereld. Het ministerie van Justitie en Veiligheid, het ministerie van Buitenlandse Zaken en de gemeente Den Haag organiseren deze high level bijeenkomst waarbij in 2022 de volgende thema's aan de orde komen: 'International Policy & Law', 'Technology & Application', 'AI & Cyber', 'Future of the City & AI'.

Hackathon for Good

De missie van Hackathon for Good is het inzetten van hackers en innovators ten behoeve van innovatieprogramma's die door overheden, bedrijven en kennisinstellingen worden opgestart en betrekking hebben op een maatschappelijk vraagstuk. Zij worden geholpen bij het maken en testen van prototype producten die gebruikmaken van data en AI. Zo is er aandacht voor het gebruik van blockchain om desinformatie tegen te gaan en AI-technologie om 'deep fakes' te herkennen. Ook werken de hackers en innovators aan bijvoorbeeld het tegengaan van voedselverspilling en overstromingen met behulp van technologie. Aan het jaarlijkse event werkten meer dan honderd deelnemers uit twintig landen aan use cases van onder andere het Openbaar Ministerie en Defensie.

2.3 BOUWSTEEN: RESEARCH EN INNOVATIE

Een netwerk van partners zorgt ervoor dat fundamenteel en toegepast onderzoek gedaan kan worden naar AI. Overheden, bedrijven, kennisinstellingen en maatschappelijke organisaties trekken samen op in onderzoek en in het ontwikkelen van AI-innovaties. Onderzoek en praktische toepassing zit dicht op elkaar, zodat ze elkaar kunnen beïnvloeden en het eindresultaat kunnen versterken.

Draagvlak

Bij innovatieve projecten is het van belang om door middel van publiek-private samenwerking de probleemeigenaar, budgethouder, verwerver en eindgebruiker te betrekken. Het is belangrijk de tijd te nemen om op bestuurlijk niveau technologieën uit te leggen en het belang ervan te onderstrepen. Ook het samenwerken met juridische afdelingen is belangrijk. Wanneer alle stakeholders vanaf het begin worden betrokken bij de projectontwikkeling, wordt voorkomen dat er onnodige en onvoorziene obstakels ontstaan voor implementatie.

Het gebruik van AI draait om maatschappelijke acceptatie. Samenwerking met burgers en maatschappelijke organisaties, bij zowel de ontwikkeling als de inzet van AI, is nodig om de afstand van de wetenschap tot de maatschappij te verkleinen.

2.3.1 HET BUNDELEN VAN KRACHTEN

Het samenwerken tussen en versterken van expertise uit verschillende disciplines is van grote meerwaarde in research en innovatie van AI.

CLAIRE en ELLIS

Het Europese research netwerk CLAIRE (Confederation of Laboratories for Artificial Intelligence Research in Europe) heeft tot doel een pan-Europees netwerk van Centers of Excellence in AI op te zetten om bestaand talent te bevorderen en een centraal punt te vormen voor de uitwisseling en interactie van onderzoekers. Het Europese ELLIS netwerk brengt daarnaast excellentie in AI-onderzoek, in het bijzonder in machine leren, samen met drie units in Nederland, namelijk de Universiteit van Amsterdam, Radboud Universiteit, en TU-Delft. Op deze manier zal de kennisstroom tussen Europese onderzoekers en hun instituten versterkt worden.

Samenwerking universiteiten Leiden, Delft en Rotterdam

Een voorbeeld op lokaal niveau is de samenwerking tussen Universiteit Leiden, de Technische Universiteit Delft en Erasmus Universiteit Rotterdam. Deze universiteiten hebben de handen ineengeslagen in de Convergentie AI, Data en Digitalisatie. Door samen te werken en kennis over het recht en bedrijfskunde te combineren met technische kennis en kennis over bestuur, beleid en ethiek vullen de expertises van de onderzoeksgroepen elkaar op een waardevolle manier aan. Met elkaar zetten de onderzoekers in op baanbrekend onderzoek gericht op de maatschappelijke uitdagingen van de toepassing van AI in het domein van vrede, recht en veiligheid. Zoals het vinden van het juist evenwicht tussen nieuwe en geoptimaliseerde functionaliteiten en het tegelijkertijd beschermen van burgers en de institutionele grondslagen van onze samenleving tegen AI-kwetsbaarheden en -afhankelijkheden.

Er liggen kansen voor multidisciplinair onderzoek op zowel toegepaste techniek-gerelateerde vragen (IN AI), zoals bijvoorbeeld NLP voor forensisch onderzoek, in toepassingen als (cyber)security en rechtspraak (MET AI), als regelgevingsvragen, ethiek, verantwoordingsdynamiek, beleidsimplementatie, transparantie uitlegbaarheid, en AI-toepassingen voor de bevordering van openbare veiligheid. De drie universiteiten hebben nauwe banden met startups, het beleid, de uitvoeringsorganen en de bedrijven die werkzaam zijn in of voor dit domein. Met deze partijen bouwen zij verder aan een toonaangevend netwerk voor AI om de kansen en uitdagingen van mensgerichte AI multidisciplinair en integraal aanpakken.

2.3.2 ETHICAL, LEGAL AND SOCIETAL ASPECTS LABS (ELSA)

ELSA staat voor 'Ethical, Legal and Societal Aspects' die een rol spelen bij de ontwikkeling en implementatie van technologische innovaties in de samenleving. In (virtuele) labs doen overheden, bedrijven, kennisinstellingen en maatschappelijke organisaties gezamenlijk onderzoek. Via co-creatie worden mensgerichte AI-oplossingen ontwikkeld, getest en geïmplementeerd. Voor het veiligheidsdomein worden de komende tijd meerdere labs uitgewerkt, waarbij de een zich specifiek focust op het defensiedomein (getrokken door TNO) en de ander zich focust op publieke AI-systemen binnen de justitiële keten (getrokken door de universiteiten in Leiden, Delft en Rotterdam). Daaromheen wordt een netwerk opgezet om te zorgen dat ideeën en resultaten onderling gedeeld en opgeschaald kunnen worden en uiteindelijk terecht komen bij de mensen die er dagelijks mee werken.

Defensie

De introductie van AI-systemen bij het ministerie van Defensie roept tal van ethische, juridische en maatschappelijke vragen op, zoals de vraag hoe systemen door mensen onder controle gehouden kunnen worden. Een andere vraag is hoe menselijke macht, waardigheid en verantwoordelijkheid behouden blijven bij het autonomie geven aan machines. Wie is verantwoordelijk voor de besluiten die machines nemen? Het ELSA Lab Defensie gaat een toekomstbestendig en onafhankelijk ecosysteem opzetten met experts en bedrijven die door Defensie geraadpleegd en ingeschakeld kunnen worden voor een verantwoord gebruik van AI binnen defensie. Ook zal een methodologie ontworpen worden die ervoor zorgt dat de ELSA aspecten altijd worden meegenomen in de ontwikkeling van AI binnen Defensie.

Justitie en Veiligheid

Steeds vaker worden AI-systemen toegepast in het publieke domein. Het aandeel van AI in democratische processen en openbare dienstverlening neemt toe, wat ernstige gevolgen kan hebben voor burgers. Dit ELSA-lab doet onderzoek naar de mogelijkheid om menselijke controle te houden over de AI-systemen die gebruikt worden door (semi-)overheden binnen het veiligheidsdomein. Het onderzoek draagt bij aan een rechtvaardige behandeling van burgers door rechtvaardige beslissingen mogelijk te maken via AI. De AI-systemen worden door een nog te ontwikkelen effectieve aanpak gecorrigeerd, zodat burgers kunnen vertrouwen op rechtsbescherming. Door diverse ELSA '*checks and balances*' in te voeren zorgt dit lab er voor dat AI-systemen in de publieke ruimte helpen de rechtsstaat te versterken.

2.3.3 NATIONAAL POLITIELAB AI

Het Nationaal Politielab Artificial Intelligence (NPAI) is een samenwerkingsinitiatief van de Nederlandse politie, de Universiteit Utrecht, de Universiteit van Amsterdam en de Technische Universiteit Delft. Het politielab maakt onderdeel uit van het nationale Innovation Center for Artifical Intelligence (ICAI). De samenwerkingspartners streven ernaar *state-of-the-art* AI-technieken te ontwikkelen om de politie te ondersteunen in haar operationele processen. Via het NPAI willen de partners de veiligheid van Nederland op een sociaal, juridisch en ethisch verantwoorde manier verbeteren. Het lab werkt aan technieken over de volle breedte van AI: *machine learning* om de juiste informatie te halen uit multimodale bronnen (foto's, tekst en video), algoritmen om te redeneren met informatie in bijvoorbeeld (juridische) documenten en opgestelde misdaadscenario's, simulaties van complexe criminele systemen en robotica. Aspecten als transparantie, privacy en uitlegbaarheid zijn hierbij net zo belangrijk als a nauwkeurigheid, berekenbaarheid en efficiëntie.

De samenwerking tussen politie en wetenschap zorgt ervoor dat de politie gebruik kan maken van de allernieuwste technieken, terwijl de wetenschap interessante problemen uit de alledaagse praktijk krijgt aangereikt om te onderzoeken. Naast wetenschappelijke artikelen heeft het lab ook al meerdere toepassingen voor de politie opgeleverd, zoals de slimme keuzehulp internetoplichting en een '*explainable AI toolbox*' voor data scientists bij de politie.

2.3.4 HET AI AND LEGAL TECH LAB (AILT)

Het AI and Legal Tech Lab van de Haagse Hogeschool heeft financiering van de Gemeente Den Haag ontvangen om de positie van Nederland als internationale hub in dit opkomende gebied te ondersteunen. Het AILT-lab voert juridische risico- en sociale impactbeoordelingen voor nieuwe technologieën om sociale gevolgen en mogelijke controverses te evalueren door de lens van juridische analyse, met behulp van studentenonderzoek. Het AILT werkt net zoals een legal clinic en heeft internationale partners op hoog niveau, waaronder de Stanford University. Het lab biedt een oefenterrein aan studenten bestuurskunde, rechten en veiligheidskunde. Door hen bloot te stellen aan echte problemen van AI-implementatie zorgt het lab ervoor dat studenten zo goed mogelijk in staat worden gesteld om in de toekomst complexe AI-governance vragen multidisciplinair op te kunnen lossen. Eerdere onderzoeksprojecten van het AILT-lab waren onder meer samenwerkingen met datawetenschappers, computerwetenschappers en juristen van gerenommeerde organisaties als CWI, Deloitte en TNO, en verschillende (inter) nationale startups.

2.4 BOUWSTEEN: STARTUPS EN SCALE-UPS

Binnen de werkgroep is een flinke groep ondernemers actief die zowel met de overheid als met kennisinstellingen projectmatig samenwerkt om hun producten continu te verbeteren en een bijdrage te leveren aan veiligheid, vrede en recht. Veel van hen zijn betrokken bij een van de projecten genoemd in deze publicatie. Naast de vele projecten geeft de werkgroep het podium aan startups en scale-ups. Zij krijgen de mogelijkheid hun innovatieve oplossingen te presenteren aan de deelnemers van de werkgroep VVR. Dit leidt tot interessante verbindingen met de deelnemers en met projecten van de werkgroep. Ook krijgen de startups en scale-ups de mogelijkheid om hun vragen en belemmeringen voor te leggen aan de deelnemers en daarover in gesprek te gaan. We noemen een paar voorbeelden van organisaties die tijdens de werkgroep bijeenkomsten voorbij zijn gekomen.

2.4.1 LEGALAIR

LegalAIR is ontstaan uit Gimix en BG Legal. Dit project heeft tot doel kennis over AI en de toepassing daarvan toegankelijk te maken voor burgers, overheden, bedrijven en kennisinstellingen. Iedereen die met AI te maken krijgt, moet eenvoudig antwoord kunnen vinden op juridische en ethische vragen. Veel AI-projecten lopen tot nu toe vertraging op of komen niet van de grond, omdat er veel kennis ontbreekt en er veel onduidelijk is. Toch zijn op veel vragen wel antwoorden te geven. Dat is het doel van dit kennisplatform. Daarnaast biedt dit platform modeldocumenten aan, zoals overeenkomsten en intellectueel eigendom clausules. Ook worden experts aan het platform gekoppeld, zodat iedereen met specifieke juridische en ethische vragen bij hen terecht kan. Uiteindelijk wil LegalAIR duidelijkheid rondom AI realiseren, zodat organisaties de mogelijkheden en kansen van AI gaan zien en benutten.

2.4.2 THE GLOBAL AI FOR GOOD COMMUNITY

De wereldwijde Global AI for Good-gemeenschap, FruitPunch AI genaamd, zet AI in voor grote uitdagingen waarvoor de mensheid staat zoals duurzaamheidsvraagstukken. Zo hielp Fruitpunch AI al eerder bij de bescherming en het behoud van het milieu en de dieren in het wild. In de toekomst wil zij een onbemand luchtvaartuig ontwikkelen dat een revolutie teweegbrengt in de manier waarop natuurreservaten worden gemonitord en geïnspecteerd. Diverse events worden door de gemeenschap georganiseerd en experts van FruitPunch AI bieden hun hulp aan bij het aangaan van maatschappelijke uitdagingen. De kracht van de Fruitpunch AI-gemeenschap wordt verbonden met de werkgroep om zo in de toekomst ook veiligheidsvraagstukken het hoofd te kunnen bieden.

2.4.3 TEGENGAAN VAN BIAS EN DISCRIMINATIE IN AI-MODELLEN

Eén van de zorgen bij het trainen van AI is dat er bias ontstaat, veroorzaakt door de gebruikte brondata. Er is sprake van een bias als externe factoren een negatieve invloed hebben op de uitkomsten. Zo kan in de brondata vastgelegd zijn dat een bepaalde bevolkingsgroep belastingfraude pleegt. Dit hoeft echter niet in het algemeen te gelden en kan het resultaat zijn van onderzoek in een bepaalde periode. Een AI-systeem dat deze data krijgt om te verwerken of om mee te trainen, kan mogelijk de bias in de data verder versterken door bijvoorbeeld meer fouten te maken als het om een bepaalde bevolking gaat of sterkere negatieve beoordelingen maken over deze bevolkingsgroep.

Het is aantoonbaar mogelijk om bias en discriminatie tegen te gaan, ook al bevat de brondata vooringenomen standpunten en meningen. Centilien, een computer vision platform, heeft AI-systemen getraind en laten zien dat het tegengaan van bias en discriminatie goed mogelijk is. Het valt niet helemaal te voorkomen, maar dat is bij mensen niet anders. Wel kan het systeem inzichtelijk maken op basis waarvan het resultaat of advies tot stand is gekomen. Deze transparantie is essentieel om uiteindelijk met AI-toepassingen aan de slag te gaan.

Innovatief MKB voor Veiligheid, Vrede en Recht

Amberscript, Avantopy, awen collective, aXite Security Tools, AXVECO, Bitdefender, Centilien, Compumatica, Considerati, Cursor Asset Management, Cybermunque, Cybersprint, Data Science District, Dataworkz, dialogic, Duck Duck Goose, Dynaxion, EclecticIQ, Experience Data, FruitPunch AI, Go2the.cloud, Incontrol, Labelfuse, Oddity.ai, OPT/NET, orangehill, Pandora Intelligence, Parito, publicsonar, reprex, Researchable, robin radar systems, RosemanLabs, SemLab, Squerist, Syntho, thrive dynamic, UbiOps, Xomnia, Zeta Alpha, ZIUZ visual intelligence

3. FOCUSGEBIED: PRIVACY-BESTENDIGE INFORMATIEDELING

Eén van de focusgebieden van de werkgroep VVR is het realiseren van oplossingen voor privacy-bestendige informatiedeling tussen verschillende partijen. Er is namelijk een zeer grote hoeveelheid data beschikbaar over mensen, afkomstig van allerlei bronnen, die niet geanalyseerd mogen worden als gevolg van privacyoverwegingen. Voor verantwoord gebruik van AI is het essentieel dat data op een privacy-bestendige manier gedeeld en gebruikt kan worden, zeker in het veiligheids-domein waar privacy van burgers een nog gevoeliger onderwerp is dan in sommige andere domeinen. Daarom fungeert privacy-bestendige informatiedeling ook als bouw-steen voor de andere focusgebieden.

De data in het veiligheidsdomein is enorm gefragmenteerd aanwezig bij veel verschillende partijen. Gemeenten, Openbaar Ministerie, politie en veiligheidsregio's hebben bijvoorbeeld ieder relevante data voor het voorkomen van criminaliteit, opsporen en vervolgen. Om AI voor veiligheidsvraagstukken te trainen, is deze data nodig. Maar door de Algemene Verordening Gegevensbescherming (AVG) en het risico dat gegevens door een ander op straat komen te liggen, is men zeer huiverig data met elkaar te delen.

De werkgroep zoekt naar oplossingen voor dit vraagstuk, zodat data geanalyseerd en benut kan worden binnen de veiligheidsketen zonder daarbij de privacy van de burger te ondermijnen. Daarbij wordt rekening gehouden met uitlegbaarheid als AI in het publieke domein wordt gebruikt, dan moet goed uitgelegd kunnen worden hoe het werkt en hoe het tot de geleverde resultaten komt. Bekend moet zijn waarop bijvoorbeeld de adviezen van het AI-systeem zijn gebaseerd.

Het delen van data is een van de bouwstenen die cruciaal is voor een goed en verantwoord gebruik van AI. In het kader van het focusgebied 'privacy-bestendige informatiedeling' maakt de werkgroep dankbaar gebruik van de resultaten die de specifieke werkgroep Data Delen van de NL AIC oplevert. Andersom draagt de werkgroep met haar initiatieven en resultaten ook bij aan de werkgroep Data Delen.

3.1 PRIVACY ENHANCING TECHNOLOGIES

Een van de oplossingen voor de uitdagingen rond data zijn Privacy Enhancing Technologies (PET). Dit zijn cryptografische technieken om privacygevoelige gegevens, zoals persoonsgegevens, op een privacy-vriendelijke manier te verwerken. Zij bestaat uit een verzameling technieken die binnen een informatiesysteem de bescherming waarborgt van de persoonlijke levenssfeer van personen. Dat doen zij door onnodige of ongewenste verwerking van persoonsgegevens te voorkomen. PET maakt het mogelijk gegevens op zodanige wijze te gebruiken, dat alleen de minimaal vereiste informatie bekend is bij betrokkenen. Deze technologie kan ervoor zorgen dat er geen tegenstelling meer bestaat tussen privacy enerzijds en veiligheid anderzijds. De maatschappelijke opdracht om een veilige samenleving te realiseren gaat dan hand in hand met de bescherming van de persoonlijke levenssfeer.

3.1.1 MULTI-PARTY COMPUTATION

Hoe kan je als organisatie data uitwisselen zonder de privacy te schenden? Door bij het ontwerp van systemen en technologieën al direct rekening te houden met privacy. Dit wordt privacy by design genoemd. Een van die technologieën is 'multi-party computation' (MPC). Het gaat hier om een slimme manier van gezamenlijk data analyseren zonder deze te onthullen. Cryptografische technieken zorgen ervoor dat meerdere partijen gezamenlijk data analyseren en conclusies trekken zonder dat zij elkaars data ooit kunnen zien. Berekeningen en analyses worden dus uitgevoerd op versleutelde data, en alleen het eindresultaat wordt ontsleuteld. Met MPC wordt dus geen data inzichtelijk gemaakt, maar alleen conclusies op basis van die data. Bovendien zorgt deze techniek ervoor dat alleen vooraf afgesproken analyses uitgevoerd kunnen worden. Hierdoor wordt het oneigenlijk gebruik van persoonsgegevens tegengegaan. Hier wordt onder andere aan gewerkt binnen het programma Techruption in het field lab 'Multiparty Computation' op de innovatiecampus van Brightlands.

3.1.2 FEDERATED LEARNING

Een andere technologie die het stempel 'privacy by design' draagt is 'federated learning'(FL). Bij deze techniek wordt de analyse naar de data gebracht in plaats van de data naar de analyse. Met FL is het niet meer nodig om op één plek veel gevoelige data te verzamelen. De data blijft decentraal bij de data-eigenaar, terwijl het toch gebruikt kan worden voor analyses. Decentraal worden op de locaties waar zich data bevindt berekeningen gedaan. De resultaten van al deze deelberekeningen worden gedeeld met een of meerdere partijen die er totaalconclusies uit trekken. Net als MPC laat FL zien dat het delen van data niet nodig is om toch nuttige inzichten te verkrijgen uit verschillende verspreide databronnen, terwijl de privacy en vertrouwelijkheid gewaarborgd zijn.

3.1.3 MA3TCH

Ook Ma3tch is een privacy by design-technologie. Het wordt gebruikt door partijen die elkaars gegevens zouden willen gebruiken volgens de eisen van de Algemene Verordening Gegevensbescherming (AVG). Daarin staat dat partijen zo min mogelijk data mogen uitwisselen om een bepaald doel te bereiken, dat ze alléén data mogen uitwisselen voor dat ene doel (doelbinding) en dat ze daar transparant over moeten zijn. Ma3tch is een technologie die ervoor zorgt dat partijen voldoen aan de AVG bij de verwerking van persoonsgegevens. Dankzij Ma3tch is het mogelijk om zonder databestanden uit te wisselen, bij de ander te 'zien' of hij informatie heeft die relevant is. Via een autonome, anonieme analyse (a3) komt dat boven tafel, waarna gerichte informatie-uitwisseling kan plaatsvinden. Door de gerichte uitvraag van informatie, in tegenstelling tot het opvragen van grote databestanden, wordt ook de internationale samenwerking tussen landen makkelijker, en ook Nederlandse overheden en organisaties zullen meer genegen zijn om op deze manier informatie uit te wisselen.

3.1.4 SYNTHETISCHE DATA

De hierboven besproken technologieën vallen onder 'privacy enhancing technologies' (PET). Een vierde PET-technologie betreft het generen van synthetische data. Privacygevoelige informatie wordt vervangen door volledig nieuwe en kunstmatige data. Met behulp van AI is het mogelijk om op basis van het originele bestand een computersysteem alle kenmerken, relaties en statistische patronen van die data te leren, waarna deze compleet nieuwe data 'verzint' met dezelfde kenmerken, relaties en statistische patronen. Op die manier kan de data gebruikt worden voor allerlei doeleinden, zonder dat de privacy van mensen wordt geschonden. Syntho en SAS hebben in samenwerking met de NL AIC onderzoek gedaan naar de toegevoegde waarde van synthetische data en wanneer deze het best ingezet kan worden.

3.2 INITIATIEVEN EN TOEPASSINGEN

Naast te ontwikkelen technologieën zijn er verschillende initiatieven die relevant zijn voor het focusgebied 'privacy-bestendige informatiedeling'. We noemen een paar voorbeelden.

3.2.1 FIRE

In onderzoeken naar zware criminaliteit neemt de politie vaak gegevensdragers zoals mobiele telefoons, computers en harde schijven in beslag om te speuren naar belastend bewijsmateriaal. Vaak neemt de politie zoveel gegevensdragers in beslag, dat het onmogelijk is om alle data één voor één handmatig te doorzoeken. Data scientists van het NFI hebben daarom een zelflerend algoritme ontwikkeld dat specifieke afbeeldingen snel uit alle data kan pikken. Het ziet bijvoorbeeld of er wapens of drugs op een foto staan, maar herkent ook teksten op foto's, zoals kentekens of rekeningnummers op gestolen bankpassen. Deze AI-technologie is opgenomen in de zoekmachine Hansken.

3.2.2 EEN DIGITALE EN VERTROUWDE INFRASTRUCTUUR VOOR MELDINGEN

Onder de naam 'Informatiehuis Meldingen' wordt er gewerkt aan het opzetten van een digitale, vertrouwde infrastructuur (DVI). Het is een datawarehouse en kenniscentrum ineen, specifiek voor een veilige leefomgeving. Het doel van de betrokken partners (NVWA, Politie, OM Oost NL, Saxion, Samen Veilig, TU Delft, Universiteit Utrecht) is ketensamenwerking te bespoedigen, en het voorkomen en sneller detecteren van risico's, overtredingen en incidenten. Verzamelde data wordt per bronhouder (data-eigenaar) opgeslagen en via onder andere privacy enhancing technologies wordt kennis ontsloten. Om nieuwe kennis te ontsluiten worden verschillende databronnen geraadpleegd. Een voorbeeld hiervan is het samenvoegen van satelliet en droneomgevings- en openbare data om de locatie van illegale drugslaboratoria te detecteren. Kritieke randvoorwaarden voor het inrichten van een informatiehuis zijn onder andere eenduidige doelen en definities, het ontbreken van commerciële belangen, wetenschappelijke ijking en benutten van technologische innovaties, en het voldoen aan voorwaarden voor privacy en ethiek.

Een digitaal vertrouwde infrastructuur kan voor verschillende doeleinden ingezet worden. Zo wordt onderzocht of het delen van sensordata en andere relevante data via een DVI tussen bedrijven en derden leidt tot een betere datagedreven beveiliging van bedrijventerreinen. Ook wordt onderzocht of het delen van data via een DVI leidt tot waardevolle inzichten die helpen de rampenbestrijding beter in te richten en incidenten te voorkomen. Aanvullend wordt gekeken of op basis van real-time (sensor)data, afkomstig van objecten, bewoners en organisaties, een effectievere aanpak van calamiteiten mogelijk is: hulpverleners kunnen hun werk bijvoorbeeld beter en veiliger uitvoeren als ze beschikken over actuele gegevens van personen, gebouwen en installaties.

3.2.3 SUSTAINABLE RESCUE

Mensenhandel is een verborgen misdaad. Het maakt vaak verborgen onderdeel uit van criminele organisaties. Door een groot gebrek aan communicatie – het delen van data en informatie - tussen allerlei instanties die zich met het onderwerp bezighouden, is er weinig inzicht in de problematiek en weinig vooruitgang. Het probleem zit in het delen van data. Dat is bij allerlei instanties beschikbaar, maar wordt pas gedeeld als het vaak al te laat is. Daarom heeft Sustainable Rescue de FAIR (Findability, Accessibility, Interoperability and Reuse) -richtlijnen toegepast op deze data. Dat betekent dat data conform alle wet- en regelgeving aan elkaar wordt gelinkt via zogenaamde 'data stations' of 'data hotels' waarin alle instanties met elkaar worden verbonden. In samenwerking met Roseman Labs en de Data Sharing Coalition is vervolgens een prototype gerealiseerd waarmee data van informaten veilig gedeeld kan worden. Deze beveiligde informatie wordt door Deloitte gebruikt om Human Trafficking Crime Scripts te ontwikkelen die door de Politie benut kunnen worden. Omdat mensenhandel niet stopt bij de landsgrenzen kunnen deze crime scripts een sterke tool worden om kennis en inzicht ook Europees te verspreiden.

4. FOCUSGEBIED: AI, DATA EN INTELLIGENCE VOOR BESLISONDERSTEUNING

Een ***first responder,*** zoals een politieagent op straat, moet dagelijks veel besluiten nemen. Hij heeft daarom een accuraat situatiebeeld nodig waarop hij zijn beslissingen kan baseren. Door hem te ondersteunen met de analyse van real-time data en hem inzicht te geven in wat er speelt, kan hij kiezen voor de juiste interventie. Het AI-systeem dat hiervoor wordt gebruikt, zou de ***first responder*** zelfs nog een voorkeurinterventie en een alternatieve interventie kunnen aanbieden. Een ander voorbeeld van beslisondersteuning betreft rechtszaken of de meldkamer. Gesproken woorden op geluids- of beelddragers kunnen worden herkend en omgezet naar geschreven tekst en vervolgens geanalyseerd. Alle data wordt gecombineerd en de analyse leidt tot een aanbeveling aan de rechter of een actie vanuit de meldkamer. Het omzetten van woorden kan ook zijn dienst bewijzen in de rechtszaal, waar alle gesproken woorden kunnen worden omgezet naar een schriftelijk verslag van wat is gezegd. Transparantie, uitlegbaarheid, verifieerbaarheid, traceer-baarheid en onweerlegbaarheid van de aanbevelingen uit initiatieven zijn hier essentieel voor een goed en verantwoord gebruik van AI.

4.1 INITIATIEVEN EN TOEPASSINGEN

Op dit moment lopen er een aantal initiatieven en toepassingen die relevant zijn voor het focusgebied 'AI, data en intelligence voor beslisondersteuning'. We noemen een aantal voorbeelden.

4.1.1 KANSRIJKE KOPPELING

Het Centraal Orgaan opvang Asielzoekers (COA) koppelt statushouders (vluchtelingen met een verblijfsstatus) aan een gemeente. Daarbij wordt gekeken naar waar zij de beste kansen hebben om een nieuw leven op te bouwen en bij te dragen aan de maatschappij. Op basis van een digitaal profiel worden statushouders onder andere op basis van opleiding, werkervaring en ambities gematcht aan gemeentes, waar zij vervolgens huisvesting krijgen. In een project onderzoekt het COA hoe ze met behulp van AI een verbetering van de integratie uitkomsten (onder andere arbeidsmarktparticipatie en studiedeelname) kunnen realiseren.

4.1.2 QUIN

Om de leeslast van analisten in liquidatiezaken te verkleinen en onvindbare veroordeelden op te sporen, ontwikkelde TNO op basis van data en artificiële intelligentie een toepassing om het gedrag van vluchtende criminelen te voorspellen. In het programma 'Onvindbare Veroordeelden' maken Politie, het Openbaar Ministerie (OM), Centraal Justitieel Incassobureau (CJIB), de Justitiële Informatiedienst (Justid) gebruik van QUIN (Question & Investigate). De softwaretoepassing is getraind en daardoor zo slim om te voorspellen waarheen criminelen vluchten. Quin is een handig hulpmiddel om zaken sneller op te lossen waarbij tijd essentieel is maar het menselijke deel niet vervangen. Toegang tot gevoelige data in lopende onderzoeken is vaak lastig, daarom is de technologie ook getest in het televisieprogramma Hunted. Daarnaast heeft Pandora Intelligence Quin geïmplementeerd in haar scenariosoftware om zo waarschijnlijke vluchtscenario's te kunnen voorspellen bijvoorbeeld bij een plofkraak of terreurdreiging. Er zijn nog vele andere toepassingen mogelijk, waardoor deze innovatie wordt gezien als een doorbraak binnen het domein van opsporing, inlichtingen en terrorismebestrijding.

4.1.3 GERECHTELIJK EN MAATSCHAPPELIJK AANVAARDE AI IN DE RECHTSZAAL

In het project gerechtelijk en maatschappelijk aanvaarde AI in de rechtszaal zal worden onderzocht hoe verzamelde data effectief door een AI-gestuurde toolset omgezet kan worden in gerechtelijk bewijs, op een zodanige wijze dat het maatschappelijk aanvaardbaar is. Deze maatschappelijke aanvaardbaarheid wordt via proefprocessen beoordeeld. Het beoogde AI-systeem visualiseert gevonden data, ter ondersteuning van onderzoekers die zoeken naar relevant bewijs. Mens en machine komen in samenspel tot goed geïnformeerde beslissingen. Dit project levert AI- en visualisatiesoftware op, evenals een standaardrichtlijn voor digitaal bewijs en procedures voor geautomatiseerde productie van gerechtelijk bewijs. Hierin wordt samengewerkt met politierechercheurs en rechters binnen het Nationaal Politielab AI.

4.1.4 HANSKEN

De hoeveelheid te onderzoeken data en databronnen in strafzaken, voornamelijk in fraude-, moord- en seksueel misbruikzaken, neemt razendsnel toe. Om de effectiviteit en snelheid van dit onderzoek te vergroten, heeft het Nederlands Forensisch Instituut (NFI) de forensische zoekmachine Hansken ontwikkeld. Met Hansken kan snel en efficiënt worden gezocht in grote hoeveelheden in beslaggenomen gegevensdragers als computers en mobiele telefoons. Hansken maakt dit mogelijk door de gegevens geautomatiseerd te structureren en te indexeren. Op alles wat relevant kan zijn, kan worden gezocht, bijvoorbeeld op woorden en namen of eigenschappen van sporen zoals bijvoorbeeld mails, chatberichten of foto's (al dan niet gemaakt met een bepaalde camera). Dankzij AI kan vervolgens in duizenden foto's specifiek gezocht worden naar bijvoorbeeld een drugscontainer, omdat het programma met behulp van trainingsdata heeft geleerd hoe een container eruit ziet. AI maakt het ook mogelijk om bijvoorbeeld duizenden tekstberichtjes te analyseren en alleen de relevante berichtjes te tonen.

In steeds meer strafzaken wordt met Hansken gevonden sporen als bewijs gebruikt. Dit vereist gedegen forensische onderbouwing: het materiaal moet op een transparante manier verwerkt worden en de gerapporteerde sporen moeten herleidbaar zijn naar de bron (bijvoorbeeld de inbeslaggenomen telefoon of computer). Het principe van transparantie en uitlegbaarheid is hierbij enorm belangrijk. Het Nederlands Forensisch Instituut (NFI) en de Hogeschool Leiden (HS Leiden) hebben afgesproken dat studenten Informatica gaan leren gebruik te maken van Hansken. Hiervoor wordt de zoekmachine geïnstalleerd in het IoT Forensic Lab van de hogeschool op de HSD Campus. Naast dat de studenten inzicht krijgen in de werking van de zoekmachine en de AI-algoritmes, leren ze ook de zoekmachine te updaten in verband met nieuwe digitale ontwikkelingen. Door het open karakter van de software leren studenten ook zelf nieuwe tools te bouwen voor de zoekmachine en daarmee de functionaliteit uit te breiden.

4.1.5 BEELDMATERIAAL SEKSUEEL MISBRUIK MINDERJARIGEN

Honderden miljoenen afbeeldingen en video's met mogelijk seksueel kindermisbruik wordt wereldwijd naar de politie gestuurd. Het is een gigantische hoeveelheid en het kost heel veel arbeidsuren om alles uit te zoeken. Naast de afbeeldingen gaat het ook nog eens om chat logs, geografische data, gebruikersnamen, e-mailadressen, IP-adressen en geluidsopnamen. Om te voorkomen dat in die grote berg belangrijke zaken niet worden opgemerkt en blijven liggen, is AviaTor ontwikkeld door ZiuZ Visual Intelligence en Web-IQ binnen een Europees consortium met onder andere de Nederlandse en Belgische politie. Dit systeem verbetert de informatiepositie van rechercheurs, zodat zij kunnen prioriteren en de meest belangrijke zaken als eerste kunnen behandelen. Het AviaTor AI-systeem brengt grote datasets met Sexual Abuse Materials (CSAM) in kaart en koppelt daar een unieke hash-code aan. Aan de hand van deze hashes kunnen beelden opgezocht en bekeken worden, en als bewijs worden gebruikt.

4.1.6 IMPACT COALITIE SAFETY AND SECURITY-VEILIGE SMART CITIES

De Impact Coalitie Safety & Security is een samenwerking tussen gemeenten, politie, kennisinstellingen, bedrijven, VNG en HSD Office. Deze coalitie is onderdeel van de bredere Smart Society-beweging. In april 2020 werd de coalitie opgericht met een intentieovereenkomst 'Impact Coalitie Safety en Security voor Smart Society'. Binnen deze coalitie vormen de gemeenten Den Haag, Amsterdam, Almere, Eindhoven en Apeldoorn de voorhoede. De inzet van gemeenten en politie ten behoeve van publieke veiligheid wordt versterkt door de innovatieve slagkracht van deze organisaties te verhogen. Er wordt specifiek gefocust op twee toepassingsgebieden: innovaties voor crowd management en voor gebiedsbeveiliging. Mobiliteitsstromen van bezoekers krijgen daarbij extra aandacht.

4.1.7 BELEIDSIMPLEMENTATIE

AI-ondersteuning bij risico-identificatie, besluitvorming en beleidsvorming kan sterk bijdragen aan openbare en nationale (cyber)veiligheid. Er bestaan echter nog belangrijke vraagstukken als het gaat om een effectieve en verantwoorde inzet van AI. Daarom focust de Universiteit Leiden zich op de vraag hoe binnen de beleidscontext AI-gedreven analyses geïmplementeerd kunnen worden in de primaire processen van organisaties, wat dat van organisaties vraagt, en hoe dat dan tot innovaties kan leiden. Een andere vraag richt zich op het bij elkaar brengen van twee werelden binnen organisaties: enerzijds de vraag naar AI-ondersteunde inzichten en besluitvormingsondersteuning, en anderzijds het daadwerkelijke gebruik ervan. Een derde vraag heeft te maken met de verantwoording over AI en het gebruik van data op bestuurlijk niveau. De volgende vraag handelt over het inzichtelijk maken van wat algoritmes doen, en welke effecten de resultaten hebben op opeenvolgende spelers in een keten. Ook komt de vraag aan bod hoe - en door wie - normen, waarden en belangen worden afgewogen bij een verantwoord gebruik van AI.

5. FOCUSGEBIED: AI VOOR CYBERSECURITY

Ondanks toenemende investeringen in cybersecurity, kunnen de meeste organisaties de snelheid en ontwikkelingen van digitale dreigingen nauwelijks bijhouden. Een middel-grote onderneming krijgt honderdduizenden alarmmeldingen per dag. Het automatiseren van cybersecuritywerkzaamheden is dan ook een belangrijke oplossing om met beperkte menselijke en financiële capaciteiten weerbaar te worden en te blijven. Daarom ligt de aandacht binnen dit focusgebied op het ontwerpen van veilige, geautomatiseerde en privacy vriendelijke systemen en producten, en anderzijds op het ontwikkelen van technieken om systemen en producten weerbaar te maken en te houden.

Essentieel daarbij is het onderzoek naar de interactie tussen mens en geautomatiseerde AI-securitytools. Hiervoor zijn diverse kennis- en innovatievragen relevant. Namelijk, het ontwikkelen van manieren om geautomatiseerd kwetsbaarheden te signaleren in broncodes, het automatiseren van veel voorkomende stappen in het pentestproces, automatisch patchen op applicatielevel, het automatiseren van operationele taken van security-incident- en dreigingsanalyses, en het automatisch detecteren of een organisatie wel of niet compliant is ten aanzien van wetgeving, standaarden en eigen beleid.

Nieuwe innovatieve technologieën zoals de toepassing van AI en machine learning voor detectie- en responseprocessen bieden oplossingen voor het weerbaar maken en houden van organisaties. Op termijn kunnen AI en machine learning in detectie- en responseprocessen de menselijke inzet vervangen, zodat schaarse cyberprofessionals op een andere en aanvullende manier kunnen worden ingezet. Inmiddels is AI geïntegreerd in diverse processen, zoals de detectie van botnets, machine learning in software vulnerabilities en automatische detectie van digitale inbraken binnen industriële omgevingen (bijvoorbeeld in industriële ICS-SCADA systemen).

5.1 KANSEN EN MOGELIJKHEDEN

Naarmate het aantal datalekken en cyberbeveiligings-incidenten toeneemt, wordt AI steeds meer geprezen om haar nieuwe manier van automatisch detecteren van malware op een netwerk, respons begeleiden op incidenten en inbraken detecteren voordat ze zich voordoen. Het rooskleurige beeld van wat AI kan opleveren is niet helemaal verkeerd, maar de verwachtingen ten aanzien van de volgende generatie AI-technologieën moeten wel wat getemperd worden. Het zal nog jaren duren voordat AI zelfstandig en zonder menselijke inbreng tot analyses in staat is. De huidige ontwikkelingen bieden kansen om overbelaste cyberanalisten te ontzien en onderbezetting bij cyberspecialisten tegen te gaan. Ook blijkt dat het inschakelen van AI voor de verdediging tegen cybercrime aanzienlijk kosteneffectiever is ten opzichte van de situatie dat AI niet wordt ingezet. De verwachting is dat cyberaanvallen complexer, omvangrijker en dynamischer zullen worden. Het menselijk vermogen om deze aanvallen te ontdekken en daarop te anticiperen zal niet meer toereikend zijn. De ontwikkeling van nieuwe defensieve AI-systemen die met gebruikmaking van machine learning adaptief reageren, zal de cyberanalist helpen bij het verdedigen van de digitale systemen.

5.2 AUTOMATED SECURITY OPERATIONS

Security Operation Centers (SOC) worden als cruciaal beschouwd bij het detecteren van aanvallen en vormen de kern van de meeste cyberbeveiligingsstrategieën. Tegelijkertijd neemt het aantal aanvallen toe, is personeel schaars en worden cyberaanvallen in toenemende mate geautomatiseerd uitgevoerd. Dit vraagt om een upgrade van bestaande SOC-platformen. Een oplossing hiervoor is het automatiseren van security operations. In 2020 is met steun van het Ministerie van Economische Zaken en Klimaat (EZK) door TNO het consortium Automated Security Operations (ASOP) opgericht. Het consortium wil in een publiek-private samenwerking een automated security platform ontwikkelen dat organisaties in staat stelt om sneller en geautomatiseerd cyberaanvallen te ontdekken en te beantwoorden. Dit moet ervoor zorgen dat het voor de gehele keten van eindgebruikers, system integrators en ontwikkelaars beter mogelijk wordt om proactief en reactief complexe cyberaanvallen af te weren.

6. FOCUSGEBIED: INZET VAN TAAL EN SPRAAKTECHNOLOGIE

Organisaties hebben te kampen met een alsmaar groeiende hoeveelheid aan ongestructureerde data in de vorm van tekst, en tijdrovende processen zoals het maken van verslagen van vergaderingen, verhoren en andere verbale interacties. Taal is overal – en gelukkig komen er steeds meer mogelijkheden om met artificiële intelligentie tekst en spraak automatisch te analyseren. De verwachting is dat we de komende jaren met behulp van Natural Language Processing (NLP) en Automatic Speech Recognition (ASR) ons leven mooier en makkelijker kunnen maken.

Mogelijke toepassingen zijn bijvoorbeeld het vinden van relevante informatie in grote hoeveelheden tekst, het automatisch transcriberen van gesprekken tussen patiënt en dokter, het detecteren van emoties in iemands stem, het detecteren van een aantal medische condities zoals beginnende Parkinson, dementie of COVID-19, en het aansturen van apparaten door middel van automatische spraakherkenning. Al deze mogelijkheden van NLP en ASR hebben toepassingen in verschillende sectoren zoals veiligheid, overheid, onderwijs, nieuwe media, of de gezondheidszorg.

De afgelopen jaren zijn er grote ontwikkelingen geweest op het gebied van taal- en spraaktechnologie. Het blijkt echter dat deze technologie geen goede resultaten biedt voor Nederlandstalige teksten en spraak. Allereerst is de kwaliteit van generieke Nederlandstalige modellen al niet goed genoeg om direct ingezet te worden bij organisaties, en blijkt er veel behoefte te zijn aan verdere doorontwikkeling van specifieke taalmodellen die toegespitst zijn op domeinspecifieke terminologie, dialecten, straattaal of buitenlandse accenten. Deze modellen zijn echter zeer kostbaar om te ontwikkelen. Hoewel grote buitenlandse commerciële partijen zoals Google Nederlandstalige modellen aanbieden, is met name de afzetmarkt van specifieke taalmodellen te klein om deze partijen te bewegen om maatwerk te leveren. Bovendien is het voor veel organisaties onwenselijk of niet mogelijk om hun data en processen langs een niet-Nederlandse partij te sturen en afhankelijk te zijn van een groot buitenlands tech-bedrijf. Daarnaast kan niet gegarandeerd worden dat deze beschikbare taalmodellen inclusief, transparant en vrij van bias zijn.

6.1 NEDERLANDSE TAAL- EN SPRAAKTECHNOLOGIE VOOR HET VEILIGHEIDSDOMEIN

Voor opsporingstaken en allerhande taken in de strafrechtketen is het belangrijk dat er goede algoritmes zijn voor de Nederlandse taal, die ook dialecten, straattaal, accenten, kindertaal en spraakafwijkingen kunnen herkennen. Deze taal-en spraaktechnologie analyseert gesproken en geschreven Nederlandse taal, legt relaties tussen gesproken en geschreven taal, en zet gesproken woorden om naar tekst. Toepassingen als het geautomatiseerd analyseren en uitwerken van taps, rapporten, verhoren, aangiften, 112-meldingen en nog veel meer wordt hierdoor mogelijk.

Daarnaast wordt in het veiligheidsdomein spraak-technologie als veelbelovend gezien, omdat mensen ermee geïdentificeerd en geverifieerd kunnen worden. Bij identificatie wordt iemands identiteit achterhaald door zijn stemgeluid te vergelijken met een brede dataset. Verificatie werkt anders. Daarbij claimt iemand een bepaalde identiteit en stelt de software op basis van een geluidsopname vast dat een stem authentiek is en inderdaad bij die specifieke persoon (identiteit) hoort. Hiervoor is slechts een vooraf opgeslagen opname per persoon nodig.

Spraakverificatie kan een extra veiligheidsfactor bieden om online fraude tegen te gaan of gebouwen te beveiligen. In sommige gevallen kan het gebruikt worden als een handtekening. In de Verenigde Staten kunnen spraakopnames met de woorden ‘I agree’ bijvoorbeeld gebruikt worden om een digitaal contract te ondertekenen. Een stem-ID kan ook gebruikt worden voor het opsporen van criminelen. Dan gaat het om identificatie. Zo kan een geluidsfragment tijdens een winkelberoving worden beluisterd, en kan de politie de daders op basis van een brede dataset van spraak-IDs identificeren.

Daarentegen kan spraaktechnologie ook de veiligheid van mensen onder druk zetten. Spraakdata kan gestolen en misbruikt worden, bijvoorbeeld om identiteitsfraude mee te plegen. En ondanks de verbeteringen is spraaktechnologie niet foutloos en kunnen er ongelukken gebeuren. Voordat spraaktechnologie in kritieke toepassingen in de zorg, defensie, veiligheidsdomein of de maakindustrie wordt ingezet, zal de betrouwbaarheid van de technologie buiten kijf moeten staan en geïnvesteerd moeten worden in technieken die misbruik juist tegengaan. Denk bijvoorbeeld aan deep fake video’s, waarin iemands uiterlijk en stem worden nagemaakt (‘gekloond’) en die bijvoorbeeld mensen om de tuin leiden en het publieke debat ondermijnen.

6.2 NEDERLANDSE AI VOOR HET NEDERLANDS (NAIN)

Meerdere publieke en private partijen in Nederland zien kansen voor Nederlandse spraakmodellen en daarop gebaseerde AI-toepassingen. Zij willen hiermee aan de slag, maar zijn als individuele organisatie niet in staat om de taalmodellen te verbeteren, omdat ze onvoldoende kennis, trainingsdata of gespecialiseerde hardware in huis hebben, of omdat de algoritmes nog verder ontwikkeld moeten worden. Daarnaast is ook gebleken dat generieke taalmodellen niet nauwkeurig genoeg zijn om ingezet de worden in specifieke gevallen, waardoor er veel maatwerk nodig is. Voor individuele publieke partijen is het niet haalbaar om dit maatwerk zelf te ontwikkelen, en voor private partijen is het niet winstgevend om voor kleinere partijen zowel de generieke alsook de specifieke taalmodellen te verbeteren. Dit wordt marktfalen genoemd: de gevraagde technologieën kunnen niet op een winstgevende manier door de Nederlandse markt voorgebracht worden, terwijl het niet wenselijk is om afhankelijk te zijn van de grote overzeese partijen.

Daarom is het consortium Nederlandse AI voor het Nederlands (NAIN) opgericht. NAIN brengt de Nederlandse taal- en spraaktechnologie naar een hoger niveau, waarbij de ontwikkelde modellen en technologieën soeverein, inclusief, divers en transparant zijn. Om dit te realiseren wordt ingezet op de volgende punten: het verder ontwikkelen van privacy enhancing technologies om data delen van privacygevoelige taal- en spraakdata mogelijk te maken, juridische en ethische kaders voor het gebruik van NLP en ASR vast stellen en aan scherpen, het aansluiten bij of ontwikkelen van infrastructuren om de volgende generatie Nederlandse taal- en spraaktechnologie te ontsluiten.

NAIN lost het probleem op van de lage kwaliteit van Nederlandse taal- en spraaktechnologie die in zijn huidige staat niet goed tot helemaal niet inzetbaar is in veel sectoren. Het consortium is uniek vanwege zijn omvang wat betreft sectoren (meerdere sectoren zijn betrokken) en type instellingen (publiek, privaat, startups, kennis- en onderwijsinstellingen, en publieke en private organisaties in Vlaanderen). In NAIN zijn sectoren vertegenwoordigd zoals gezondheid, nieuwe media, onderwijs, commercie en veiligheid. Ook is aansluiting gezocht bij de stichting Nederlandstalige Spraak Coalitie. Gezamenlijk wordt gewerkt aan het bij elkaar brengen van lopende initiatieven (voorkomen dubbel werk, efficiënt benutten van projectkosten), kennisopbouw taal- en spraaktechnologie, soevereiniteit (meer controle en onafhankelijkheid ten opzichte van grote buitenlandse commerciële partijen) en inclusie (betere prestaties van taal- en spraaktechnologie als het gaat om dialecten, accenten en straattaal). Uiteindelijk moet dit leiden tot zowel een generiek als meerdere specifieke Nederlandse taalmodellen van hoge kwaliteit, verbeterde algoritmes (verwerkt in software), een licentiemodel of andere vorm van ontsluiting van de ontwikkelde taalmodellen, en juridische en ethische kaders voor gebruik van taal- en spraaktechnologie in Nederland.

Toepassingen als het geautomatiseerd uitwerken van taps, rapporten, verhoren, aangiften, 112-meldingen en nog veel meer wordt nu mogelijk. Een vooruitzicht voor toepassing in de mediasector is bijvoorbeeld dat met spraaktechnologie geautomatiseerde ondertitels onder films kunnen worden gezet. In de zorg doemt het perspectief op dat zorgverleners tijdens hun werk mondeling rapporteren wat ze doen, waarna dat automatisch wordt verwerkt en opgeslagen als administratieve rapportage en verantwoording. Binnen de NL AIC wordt voor cross-sectorale toepassingen nauw samengewerkt met de andere sectorale werkgroepen.

In de landschapskaart van november 2021 presenteert het NAIN-consortium met steun van de Zuid-Holland AI hub de huidige staat van taal- en spraaktechnologie in Nederland en Vlaanderen. Vanuit deze landschapskaart kan de komende vijf jaar verder worden gewerkt aan de ontwikkeling van state-of-the-art soevereine Nederlandstalige taal- en spraaktechnologie, die inclusief, divers, transparant en uitlegbaar is, en waar domein specifieke extensies aan gekoppeld kunnen worden. De uiteindelijke resultaten van dit project zijn overal in de Nederlandse samenleving bruikbaar.

NEEM CONTACT MET ONS OP

Als werkgroep waarborgen we een mensgerichte aanpak waarbij AI-toepassingen op een effectieve, veilige en verantwoorde manier wordt ontwikkeld én ingezet. We kijken vooruit. We verbinden vanuit de inhoud, borgen geleerde lessen en zorgen waar mogelijk voor opschaling. We jagen investeringen aan, ondersteunen kansrijke projecten en laten deze landen bij de eindgebruikers. We zetten de burger centraal, creëren draagvlak voor innovatieve oplossingen en houden de veiligheidsprofessional zowel op bestuurlijk als operationeel niveau betrokken.

Uw bijdrage is hierin noodzakelijk. Met welke thema's die hier beschreven staan, bent u al bezig? Door welke nieuwe thema's bent u geraakt? Waar zou u willen aanhaken? Welke projecten zou u willen opstarten die helpen om de ambities en doelstellingen in dit document te realiseren? Met wie zou u hierin willen optrekken?

Wij horen het graag van u. Neem daarom contact met ons op! Wij denken graag met u mee over projectontwikkeling, consortiumvorming en financieringsmogelijkheden.

E-mail: vredeveiligheidenrecht@nlaic.com

DEELNEMERS

Aan deze publicatie hebben meer dan honderd vertegenwoordigers van bedrijfsleven, kennisinstellingen, maatschappelijke organisaties en de overheid meegewerkt. Speciale dank gaat uit naar de kerngroep.

KERNGROEP

Avans Hogeschool
Ben Kokkeler

Centraal Orgaan Opvang Asielzoekers
Sjef van Grinsven

Centilien
Gerard Kanters

Data Science District
Kai Lemkes

Dienst Justitiele Inrichtingen
Ramona Apostel

Dynaxion
Cor Datema

Erasmus Universiteit
Klaus Heine

Faculty of Impact
Frans Nauta

Haagse Hogeschool
Liduine Bremer
Elif Kiesow Cortez

Heijnen Consulting/Samen Veiliger
Alexander Heijnen

Hogeschool Leiden
Jos Griffioen
Hans Henseler

Jheronimus Academy of Data Science (JADS)
Peter de Kock (Pandora Intelligence)
Liesbeth Leijssen

Justitiële Informatiedienst
Tom Schepers

Ministerie van Justitie en Veiligheid
Ron Hanoeman
Bas ter Luun
Michel van Leeuwen
Caspar Heetman
Jitske Wuite

NFI
Lisanne van Dijk
Erwin van Eijk

Openbaar Ministerie (OM)
Tjistke Visser

Politie
Theo van der Plas
Bas Testerink

Radboud Universiteit
Frederik Zuiderveen Borgesius

Reclassering Nederland
René Poort

Researchable
Eduard van Pagee

Rijksuniversiteit Groningen
Bart Verheij

Saxion Hogeschool
Remco Spithoven

Security Delta HSD
Marlou Snelders
Joris den Bruinen

Sustainable Rescue
Paul Fockens

TNO
Saskia Lensink
Joachim de Greeff
Eelko Steenhuis

TU Delft
Marlou Smulders
Eveline Vreede

Universiteit van Amsterdam
Marcel Worring (Nationaal Politielab AI/ICAI)

Universiteit Leiden
Bram Klievink

Universiteit Utrecht
Floris Bex (Nationaal Politielab AI/ICAI)

Zuyd Hogeschool
Mark Liedekerken

ZiuZ
Jos Flury

INITIATIEFNEMERS

HSD securitydelta.nl

Ministerie van Justitie en Veiligheid

NL AI Coalitie

dutch digital delta connect and create

KADERS VOOR HET GEBRUIK VAN AI

		Doel	Karakter	Vindplaats en opleverdatum
Normen kaders	Recommendation on the Ethics of AI – UNESCO	Aanbevelingen ten behoeve van landen voor wetgeving en beleid om AI ethisch in te zetten.	Niet bindend: Aanbeveling	Het voorstel voor de aanbevelingen wordt eind 2021 verwacht. De preliminary study is reeds op de UNESCO-site te vinden.
	E thics Guidelines for Trustworthy AI – Europese Commissie	Ethische richtsnoeren ontwikkelt op verzoek van de Europese Commissie die organisaties kunnen gebruiken wanneer zij AI inzetten.	Niet bindend: Hulpmiddel	De richtsnoeren zijn in april 2019 opgeleverd.
	Juridisch raamwerk voor AI (CAHAI) – Raad van Europa	Met een feasibility study is de noodzaak bepaald om te komen tot een juridisch raamwerk (bijvoorbeeld in de vorm van een verdrag) ter verdere bescherming van de democratie, rechtsstaat en fundamentele rechten.	N.t.b. mogelijk bindende en niet bindende onderdelen	De feasibility study is in december 2020 afgerond en door de CAHAI (tijdens de plenaire vergadering) aangenomen.
	Wetgeving m.b.t. algoritmen en AI – Europese Commissie	De Europese Commissie publiceerde in april 2020 haar Witboek AI dat als doel had inzicht te geven in de denkrichtingen m.b.t. mogelijke wetgeving op het gebied van algoritmen en artificiële intelligentie. In opvolging hiervan heeft de Commissie op 21 april 2021 een voorstel voor een verordening om het juridisch kader verder aan te vullen gepubliceerd.	N.t.b. mogelijk bindende en niet bindende onderdelen	Het kabinet heeft uw Kamer in mei 2020 haar appreciatie van het Witboek AI toegezonden. 21 april 2021 is het voorstel voor een verordening AI gepubliceerd.
	Richtlijnen voor het toepassen van dataanalyse door de overheid – het kabinet	Overheidsorganisaties handvatten te bieden om op rechtmatige wijze (algoritmische) data-analyse toe te passen.	Niet bindend: Hulpmiddel	Worden doorlopend aangescherpt. In deze brief wordt de verwezen naar de publicatie van de laatste versie.
Toezicht kaders	Impact Assessment voor Mensenrechten bij de inzet van Algoritmen – het kabinet	Instrument dat overheidsorganisaties helpt om in de gehele levenscyclus van technologische systemen risico's voor mensenrechtenschendingen in kaart te brengen en te mitigeren.	Niet bindend: Hulpmiddel	Het IAMA wordt naar verwachting eind juni van 2021 opgeleverd.
	Data Protection Impact Assessment (DPIA)	In kaart brengen van risico's voor rechten en vrijheden van personen bij het verwerken van persoonsgegevens.	Verplicht voor hoog-risico verwerkingen (artikel 35 AVG)	Artikel 35 AVG.

		Doel	Karakter	Vindplaats en opleverdatum
Toezicht kaders	T oetsingskader Algoritmen – ARK	Het toetsingskader is een praktisch instrument dat de ARK en andere overheidsorganisaties kunnen gebruiken om te toetsen of algoritmes aan kwaliteitscriteria voldoen én of de risico's in beeld zijn en worden beperkt.	Niet bindend.	Het toetsingskader is op de site van de ARK te vinden.
	Normenkader Algoritmen – ADR	Een door de ADR ontwikkeld toetsingskader dat op termijn wellicht ingezet kan worden voor het structureel auditen van algoritmes bij de overheid.	N.t.b.	Het kader wordt in 2021 a.d.h.v. specifieke casuïstiek getoetst en doorontwikkeld.
Handreikingen en methodieken	AI-systeemprincipes non-discriminatie – het kabinet	De AI-systeemprincipes voor non-discriminatie zijn in de vorm van een handreiking een hulpmiddel voor overheidsorganisaties om ontwikkelaars van AI-systemen helpt discriminatie al in de ontwikkelfase van een AI-systeem te ondervangen.	Niet bindend: handreiking.	Deze handreiking is in januari 2021 opgeleverd en wordt met deze brief meegestuurd.
	Code Goed Digitaal Openbaar Bestuur – het kabinet	Addendum op Code Goed Openbaar Bestuur specifiek voor digitalisering.	Niet bindend.	De code wordt met deze brief meegestuurd.
	Calculemus-FLINT – het kabinet	Systematiek om transparantie van gemaakte keuzes in algoritmen te vergroten. Om wet- en regelgeving transparanter te maken worden de juridische bronnen waar publieke diensten op gebaseerd zijn, expliciet gemaakt.	Niet bindend: hulpmiddel.	Wordt verder uitgewerkt a.d.h.v casuïstiek.
	Artificial Intelligence Impact Assessment - ECP	De (AIIA) helpt bedrijven artificiële intelligentie verantwoord in te zetten door een kader te bieden waarmee een bedrijf kan toetsen wat de impact van het algoritme op mensen is.	Niet bindend: hulpmiddel.	Het AIIA is te vinden op de site van het ECP.
	Algoritmeregister (of alternatief instrumentmet vergelijkbare functie) in ontwikkeling – het kabinet	Het bieden van een standaardformat om overzicht te houden – en transparant te zijn over- de inzet van algoritmen.	N.t.b.	Het kabinet zal een proef doen met een concreet transparantie-instrument, zoals een algoritmeregister.
	Code Kinderrechten Online – het kabinet	Het bieden van handvatten aan overheden en bedrijven om de rechten van kinderen online te borgen.	Niet bindend	Het bieden van handvatten om de rechten van kinderen online te borgen

Eindredactie
Marlou Snelders (Security Delta HSD, Nederlandse AI Coalitie)
Martin Bobeldijk (Turnaround Communicatie)

Contact:
Email — vredeveiligheidenrecht@nlaic.com
Website — nlaic.com

Januari 2022