ARTIFICIËLE INTELLIGENTIE, MENSENRECHTEN, DEMOCRATIE, EN DE RECHTSSTAAT

EEN INTRODUCTIE

VOORBEREID TER ONDERSTEUNING VAN DE HAALBAARHEIDSSTUDIE AANGENOMEN DOOR HET AD HOC COMITÉ INZAKE ARTIFICIËLE INTELLIGENTIE VAN DE RAAD VAN EUROPA

DAVID LESLIE, CHRISTOPHER BURR, MHAIRI AITKEN, JOSH COWLS, MIKE KATELL, & MORGAN BRIGGS

Met een voorwoord door

LORD TIM CLEMENT-JONES

Nederlandse vertaling: Peggy Valcke, Catelijne Muller, Bram Goetry

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

The Alan Turing Institute

The Alan Turing Institute, in samenwerking met het Vlaams Kenniscentrum Data & Maatschappij en ALLAI

Het Openbare Beleidsprogramma van “The Alan Turing Institute” werd opgericht in mei 2018 met als doel het ontwikkelen van onderzoek, middelen, en technieken om overheden bij te staan bij hun innovatie op het vlak van data-intensieve technologieën gericht op de verbetering van de levenskwaliteit. Wij werken samen met beleidsmakers om te onderzoeken hoe data-wetenschap en artificiële intelligentie het openbaar beleid kunnen onderbouwen en de openbare dienstverlening kunnen verbeteren. Wij zijn van mening dat regeringen de voordelen van deze technologieën alleen zullen kunnen ervaren indien zij ethische en veiligheidsoverwegingen tot een eerste prioriteit maken.

Over het KDM – Het Kenniscentrum Data & Maatschappij is een samenwerking tussen drie universitaire onderzoeksgroepen: imec-SMIT-VUB, KU-Leuven CiTiP en imec-MICT-UGent. Het maakt deel uit van het Vlaams Beleidsplan Artificiële Intelligentie en krijgt steun van de Vlaamse overheid (EWI). Het KDM is de centrale hub voor de juridische, maatschappelijke en ethische aspecten van data-gedreven applicaties en AI-toepassingen. Het draagt via diverse activiteiten bij aan het debat en het creëren van een maatschappelijk draagvlak voor AI en data-gedreven toepassingen (https://data-en-maatschappij.ai/).

Over ALLAI - ALLAI is een onafhankelijke organisatie die zich richt op verantwoord(e) ontwikkeling, inzet en gebruik van AI. ALLAI is opgericht door de 3 Nederlandse leden van de High Level Expert Group on AI van de Europese Commissie. ALLAI richt zich op de brede maatschappelijke impact van AI vanuit ethisch, juridisch en sociaal-maatschappelijk oogpunt. Door actieve betrokkenheid bij AI-beleidsvorming, het vergroten van kennis en bewustwording over de kansen en mogelijkheden van AI, het vertalen van AI-beleid naar de maatschappij en het bevorderen van een inclusief AI-debat zorgt ALLAI dat AI op een verantwoorde manier in onze samenleving landt. https://www.allai.nl

Merk op: deze brochure is een levend document dat verder zal evolueren en verbeteren dankzij input van lezers, belanghebbenden en geïnteresseerde partijen. Wij hebben uw medewerking nodig. Deel uw feedback met ons via policy@turing.ac.uk (in het Engels). Dit onderzoek werd mede mogelijk gemaakt door een subsidie van ESRC (ES/T007354/1) en uit publieke fondsen die het Openbare Beleidsprogramma van “The Alan Turing Institute” mogelijk maken.
https://www.turing.ac.uk/research/research-programmes/public-policy

De in dit werk weergegeven meningen vallen onder de verantwoordelijkheid van de auteurs en geven niet noodzakelijk het officiële beleid van de Raad van Europa weer. De informatie waarop een groot deel van deze brochure is gebaseerd, werd overgenomen uit de Haalbaarheidsstudie aangenomen door het Ad Hoc Comité inzake Artificiële Intelligentie en gepubliceerd in december 2020. Lezers wordt aanbevolen de Haalbaarheidsstudie te raadplegen voor meer details: https://rm.coe.int/cahai-2020-23-finaleng-feasibility-study-/1680a0c6da

---

Citeer dit werk als:

Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., en Briggs, M. (2021). Artificiële intelligentie, mensenrechten, democratie, en de rechtsstaat: Een introductie. The Alan Turing Institute - i.s.m. KDM en ALLAI, Nederlandse vertaling door Valcke, P., Muller, C. en Goetry, B.

INHOUDSOPGAVE

01 INLEIDING **5**

02 HOE WERKEN AI SYSTEMEN? **7**

Technische concepten

Soorten machinaal leren

Fases in de levenscyclus van AI

03 MENSENRECHTEN, DEMOCRATIE EN DE RECHTSSTAAT **13**

De onderlinge afhankelijkheid van Mensenrechten, Democratie en de Rechtsstaat

04 KANSEN EN RISICO'S VAN AI/ML EN HUN IMPACT OP MENSENRECHTEN, DEMOCRATIE EN DE RECHTSSTAAT **15**

05 BEGINSELEN EN PRIORITEITEN VOOR EEN JURIDISCH KADER **18**

Beginselen, Rechten en Plichten met elkaar verbinden

Aanvullende Overwegingen

06 LANDSCHAP VAN JURIDISCHE INSTRUMENTEN **25**

Internationale Wetgevende Kaders

Huidige “Soft Law” Benaderingen

Nationale Wetgevende Instrumenten

De Rol van Private Actoren

Huidige Beperkingen

Toekomstige Behoeften en Mogelijkheden

Opties voor een Juridisch Kader

07 PRAKTISCHE MECHANISMEN TER ONDERSTEUNING VAN HET JURIDISCHE KADER **32**

De Rol van Handhavingsmechanismen

De Rol van de Verschillende Actoren

Voorbeelden van Types Handhavingsmechanismen

Follow-up Mechanismen

08 CONCLUSIE **36**

09 BIJLAGEN **37**

Glossarium

Werkzaamheden van de Raad van Europa en anderen op het vlak van AI en aangrenzende gebieden tot op heden

VOORWOORD

Het is nog nooit zo duidelijk geweest, zeker nu het voorbije jaar COVID onze afhankelijkheid van digitale technologie heeft aangetoond, dat het vertrouwen van het publiek in de adoptie van AI behouden moet blijven.

Daartoe moeten we, terwijl we de kansenvan AI ten volle benutten, de risico's beperken die aan de toepassing ervan verbonden zijn. Dit brengt de noodzaak mee van duidelijke normen voor toerekenbaarheid en ethisch gedrag.

Was 2019 het jaar waarin landen hun handtekening zetten onder internationaal overeengekomen ethische beginselen voor AI, zoals die in de OESO-Aanbeveling over AI, en de niet-bindende beginselen van de G20 over AI, dan markeerde 2020 het moment waarop de internationale AI-gemeenschap begon te bepalen hoe deze beginselen in de ontwikkeling en toepassing van AI-systemen konden worden ingebouwd.

Om van ethische AI een realiteit te maken, moeten de risico's van AI in hun context worden beoordeeld, in het bijzonder wat betreft de impact op burgerlijke en sociale rechten. Vervolgens, afhankelijk van het beoordeelde risico, dienen normen te worden vastgesteld of dient regulering te worden ingevoerd voor het ethisch ontwerpen, ontwikkelen en inzetten van AI-systemen.

Een belangrijk initiatief in dit proces is de Haalbaarheidsstudie die in december 2020 werd goedgekeurd door het Ad Hoc Comité voor Artificiële Intelligentie van de Raad van Europa (CAHAI). Daarin worden de mogelijkheden verkend voor een internationaal juridisch kader gebaseerd op de rechtsnormen van de Raad van Europa inzake mensenrechten, democratie en de rechtsstaat.

De kernvraag is of er oplossingen zijn voor de specifieke risico's en mogelijkheden die AI-systemen meebrengen, die kunnen en moeten worden opgenomen in bindende en niet-bindende internationale juridische instrumenten op het niveau van de Raad van Europa, hoeder van het Europees Verdrag tot Bescherming van de Rechten van de Mens, Verdrag 108+ inzake de verwerking van persoonsgegevens en het Europees Sociaal Handvest.

De Raad en CAHAI staan op het punt een publiek consultatie over de Haalbaarheidsstudie te lanceren. Daarom is het cruciaal dat de maatschappelijke en regelgevende implicaties van de daarin voorgestelde op beginselen gebaseerde aanpak ten volle worden begrepen, wil men het potentieel van AI verwezenlijken en de juiste keuzes maken, in het bijzonder wat betreft juridische instrumenten en de toezichts- en handhavingsmechanismen.

Deze prachtige brochure opgesteld door het Alan Turing Institute [N.B. en vertaald naar het Nederlands door het KDM en ALLAI] vormt een uitstekende gids bij de Haalbaarheidsstudie; op een heldere manier schetst hij de context ervan en vergroot hij de inhoudelijke toegankelijkheid. De brochure zal ongetwijfeld de publieke betrokkenheid vergroten en bijdragen tot een breed, en tegelijk geïnformeerd, debat. Dit is een vitaal onderdeel van het openbaar beleid waar een brede en geïnformeerde discussie door velen, met name over de te beschermen waarden, van cruciaal belang is. Dankzij deze brochure zal deze besluitvorming niet uitsluitend overgelaten worden aan een klein aantal specialisten.

Lord Tim Clement-Jones
Londen, 2021

01 INLEIDING

HET DOEL VAN DEZE BROCHURE

Het is een opmerkelijk feit dat door de snelle vooruitgang op het gebied van artificiële intelligentie (AI) en datagestuurde technologieën in de voorbije twee decennia de hedendaagse samenleving op een keerpunt staat en de vraag over hoe de toekomst van de mensheid eruit zal zien zich opdringt. Enerzijds belooft de opkomst van maatschappelijk nuttige AI-innovatie ons onder meer te helpen bij de aanpak van klimaatverandering en het verlies aan biodiversiteit; de verbetering van de medische zorg, de levensstandaard, het vervoer en de landbouwproductie (mits op een billijke wijze); en het aankaarten van sociale onrechtvaardigheden en materiële ongelijkheden waarmee de wereld vandaag te kampen heeft. Anderzijds signaleert de proliferatie van onverantwoorde AI-innovaties problemen die potentieel kunnen opduiken als de technologische vooruitgang op de huidige verontrustende koers verder gaat.

Alarmsignalen zien we bijvoorbeeld bij de toenemende risico's voor onze gekoesterde rechten op privacy, zelfexpressie, vereniging en instemming, alsook voor andere burgerlijke en sociale vrijheden, die digitale controle-infrastructuren zoals live gezichtsherkenning in toenemende mate met zich meebrengen. Ongerustheid heerst ook over de zichtbaar wordende transformerende effecten van grootschalige proliferatie van individuele targeting, algoritmische selectie en datagestuurde gedragsmanipulatie, die de inkomsten van Big Tech-platformen omhoog doen schieten, terwijl ze tegelijkertijd wereldwijde crisissen van sociaal wantrouwen, besmettingen met desinformatie en toenemende niveaus van culturele en politieke polarisatie in de hand werken. We zien ook alarmsignalen in de manier waarop de toepassing van voorspellende risicomodellen en algoritmisch versterkte digitale opsporingscapaciteiten, en dit op gebieden met een grote impact, zoals rechtshandhaving, patronen van structurele discriminatie, systemische marginalisering en ongelijkheid versterken en verder verankeren.

Het Comité van Ministers van de Raad van Europa erkent de noodzaak van democratisch gestuurd menselijk ingrijpen om AI-innovatie op het juiste spoor te zetten, en heeft daarom in september 2019 de taakomschrijving vastgesteld voor het Ad Hoc Comité voor Artificiële Intelligentie (CAHAI). CAHAI heeft als opdracht de haalbaarheid en de mogelijke elementen van een juridisch kader voor het ontwerpen, ontwikkelen en gebruik van AI-systemen te onderzoeken, dat in overeenstemming is met de normen van de Raad van Europa op de onderling samenhangende gebieden van de mensenrechten, democratie en de rechtsstaat.

Als eerste en noodzakelijke stap in de uitvoering van deze verantwoordelijkheid heeft CAHAI een *Haalbaarheidsstudie* aangenomen in plenaire vergadering in december 2020. Daarin verkent het opties voor een internationaal juridisch kader dat bestaande lacunes in de wetgeving opvult en het gebruik van bindende en niet-bindende rechtsinstrumenten afstemt op de specifieke risico's en kansen die AI-systemen met zich meebrengen. In de *Haalbaarheidsstudie* wordt onderzocht hoe de fundamentele rechten en vrijheden die reeds in de internationale mensenrechtenwetgeving zijn gecodificeerd, kunnen worden gebruikt als basis voor een dergelijk wetgevend kader. De studie stelt negen beginselen en prioriteiten voor die afgestemd zijn op de nieuwe uitdagingen ten gevolge van het ontwerpen, het ontwikkelen en het gebruik van AI-systemen. Wanneer deze beginselen en prioriteiten in wetgeving worden gecodificeerd, ontstaat een reeks samenhangende rechten en verplichtingen die ervoor moeten zorgen dat het ontwerp en het gebruik van AI-technologieën in overeenstemming is met de waarden van de mensenrechten, democratie en de rechtsstaat. De *Haalbaarheidsstudie* concludeert dat de huidige regels en rechtsstelsels noch toereikend zijn om deze fundamentele waarden te waarborgen in de context van AI, noch op zichzelf geschikt zijn om een innovatieve omgeving voor AI te creëren die voldoende betrouwbaar kan worden geacht om AI en data-intensieve technologieën in de juiste richting te sturen. Een nieuw juridisch kader is vereist.

Opzet van deze brochure is om de belangrijkste concepten en beginselen uit de *Haalbaarheidsstudie* van CAHAI bij een algemeen, niet-technisch publiek te introduceren. Ook beoogt hij enige achtergrondinformatie te verschaffen over AI-innovatie, de wetgeving inzake mensenrechten, technologiebeleid, en handhavingsmechanismen die in de studie aan bod komen. In overeenstemming met het streven van de Raad van Europa naar breed overleg, bereik en betrokkenheid van meerdere belanghebbenden, is deze brochure ontworpen om een zinvolle en geïnformeerde deelname van een inclusieve groep belanghebbenden te vergemakkelijken, nu CAHAI feedback en advies wenst te krijgen over de essentiële vraagstukken die in de *Haalbaarheidsstudie* naar boven zijn gekomen.

HOE DEZE BROCHURE TE GEBRUIKEN

Deze brochure is ontwikkeld voor zowel lezers zonder technische achtergrond als lezers met een technische achtergrond die hun kennis over een of meer van de onderwerpen uit de *Haalbaarheidsstudie* willen bijspijkeren. De hoofdstukken zijn op modulaire wijze geschreven, wat betekent dat de lezer de mogelijkheid heeft die onderdelen te selecteren die hem of haar het meest interesseren (om zich daarop te concentreren), dan wel om deze brochure van begin tot eind door te nemen.

De eerste drie hoofdstukken bieden belangrijke achtergrondinformatie over AI en machinaal leren (Hoofdstuk 2); mensenrechten, democratie en de rechtsstaat (Hoofdstuk 3); en de risico's en mogelijkheden die AI-systemen bieden in de context van de mensenrechten (Hoofdstuk 4). Vervolgens bespreekt deze brochure enkele van de meer specifieke onderwerpen die in de *Haalbaarheidsstudie* aan bod komen. Hoofdstuk 5 licht de negen beginselen en prioriteiten toe die door CAHAI als anker zijn voorgesteld voor een op waarden gebaseerd en horizontaal (sectoroverschrijdend) juridisch kader. Vervolgens worden de raakpunten gepresenteerd tussen deze beginselen en prioriteiten, enerzijds, en de belangrijkste rechten en plichten aan de hand waarvan deze in wetgeving kunnen worden omgezet anderzijds. Hoofdstuk 6 biedt een overzicht van mogelijke juridische instrumenten die geïntegreerd kunnen worden in een groter geheel van bindende en niet-bindende juridische mechanismen. Ten slotte presenteert hoofdstuk 7 het spectrum van handhavingsinstrumenten die beschikbaar zijn om de vereisten opgelegd door een juridisch kader, te ondersteunen, te verwezenlijken en te onderbouwen.

Aan het einde van deze brochure vindt u een verklarende woordenlijst van relevante termen en een geannoteerde lijst van publicaties, waaronder enkele van de eerdere werkzaamheden van de Raad van Europa en anderen op het vlak van AI-normen en -regulering en aangrenzende gebieden van technologisch beleid.

Omdat niets kan tippen aan het origineel, raden wij de lezers ten zeerste aan zich verder te verdiepen in de *Haalbaarheidsstudie* zelf en deze brochure louter te gebruiken als een hulpmiddel, voor contextuele informatie, verduidelijking en een beknopte presentatie.

02 HOE WERKEN AI SYSTEMEN?

Alvorens te onderzoeken hoe een kader van bindende en niet-bindende juridische instrumenten het ontwerp, de ontwikkeling en het gebruik van AI-technologieën kan afstemmen op mensenrechten, democratie en de rechtsstaat, geven we eerst een uiteenzetting van technische basisconcepten, types machinaal leren en de stadia van de AI-levenscyclus.

TECHNISCHE CONCEPTEN

PERSOONSGEGEVENS

Gegevens die kunnen worden gebruikt om een persoon te identificeren. Voorbeelden van persoonsgegevens zijn onder meer voor- en achternaam, adres, locatiegegevens en vormen van identificatie (bv. paspoort, nationale identiteitskaart).

ALGORITME

Een rekenkundig proces of een reeks regels die worden uitgevoerd om een bepaald probleem op te lossen. Meestal wordt een computer gebruikt om complexe algoritmes uit te voeren, maar een mens kan dit evenzeer, bijvoorbeeld door een recept te volgen of een wiskundige formule te gebruiken om een vergelijking op te lossen.

MACHINAAL LEREN (ML)

Een computertechniek die gebruikt wordt om patronen in data te vinden en voorspellingen te doen over de uitkomst van een bepaald geval. "Leren" is licht misleidend, aangezien de computer niet op dezelfde manier leert als de mens. In plaats daarvan is de computer in staat overeenkomsten en verschillen in de data te vinden door het herhaaldelijk aanpassen van de parameters (vaak "training" genoemd). Wanneer de ingevoerde data veranderen, veranderen ook de resultaten, wat betekent dat de computer nieuwe patronen leert detecteren. Dit wordt bereikt door een wiskundige formule toe te passen op grote hoeveelheden ingevoerde data om een overeenkomstig resultaat te verkrijgen. Dit wordt in meer detail beschreven in de volgende sectie.

ARTIFICIËLE INTELLIGENTIE (AI)

In de afgelopen decennia is AI op vele manieren gedefinieerd, maar voor deze brochure houden we het bij een definitie die beschrijft wat het doet, d.w.z. welke rol het speelt in de menselijke wereld: AI-systemen zijn algoritmische modellen die cognitieve of waarnemende functies in de wereld uitvoeren die voorheen voorbehouden waren aan denkende, beoordelende en redenerende mensen.

BIG DATA

Datasets of gegevenssets die omvangrijk zijn, vaak grote hoeveelheden opslag vereisen, en enorme hoeveelheden kwantitatieve gegevens bevatten die gebruikt kunnen worden om patronen of trends aan het licht te brengen. De data die in deze grote datasets zijn opgenomen kunnen uiteenlopen qua type (bv. getallen, woorden, beelden) en kunnen ofwel specifiek dienen voor een bepaald doel en in een tabelvorm zitten (gestructureerd), of algemeen en gevarieerd (ongestructureerd) zijn.

DATAWETENSCHAP

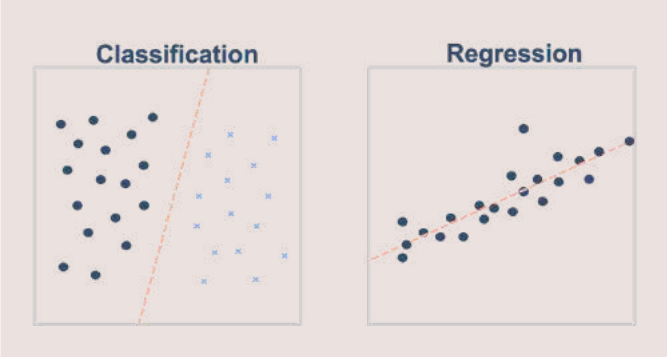
Een gebied dat elementen uit verschillende disciplines omvat, waaronder computerwetenschap, wiskunde, statistiek en sociale wetenschappen, en dat in het algemeen gericht is op het extraheren van inzichten en patronen uit datasets om een specifieke vraag te beantwoorden of een specifiek probleem op te lossen.

INTERPRETEERBAARHEID

Als een mens kan identificeren hoe een AI- of machinaal-leersysteem tot een bepaalde beslissing is gekomen, of kan verklaren waarom het zich op een bepaalde manier heeft gedragen, dan kan het systeem als interpreteerbaar omschreven worden. Interpreteerbaarheid kan ook betrekking hebben op de transparantie van de processen volgens welke het systeem ontwikkeld is.

SOORTEN MACHINAAL LEREN

GESUPERVISEERD LEREN



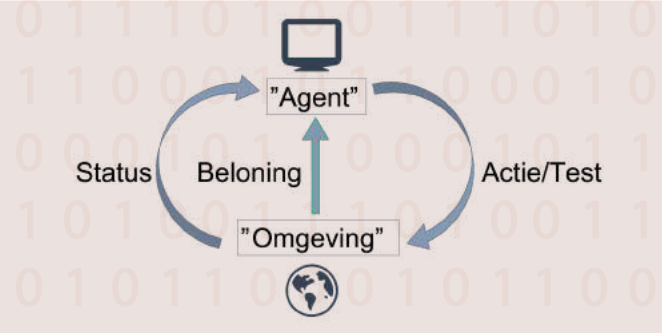
Modellen voor gesuperviseerd leren (supervised learning) worden getraind op datasets die gelabelde data bevatten. Deze modellen "leren" aan de hand van talrijke voorbeelden die worden gebruikt om het algoritme te trainen bij het afstemmen van inputvariabelen, i.e. de ingevoerde data (vaak kenmerken genoemd), op gewenste outputs (ook doelvariabelen of labels genoemd).

Op basis van deze voorbeelden zijn ML-modellen in staat patronen te identificeren die input- of invoerdata aan outputdata of resultaten koppelen. Dergelijke ML-modellen kunnen deze patronen vervolgens reproduceren door gebruik te maken van de tijdens de training verfijnde regels om op die manier nieuwe inputdata te transformeren in classificaties of voorspellingen. Een klassiek voorbeeld van gesuperviseerd leren is het gebruik van verschillende variabelen, zoals de aanwezigheid van woorden als "loterij" of "je hebt gewonnen", om te voorspellen of een e-mail al dan niet als spam moet worden geclassificeerd. Gesuperviseerd leren kan de vorm aannemen van een classificatie, zoals de voorspelling dat een e-mail al dan niet spam is, of van regressie, waarbij de relatie tussen inputvariabelen en een doelvariabele wordt bepaald. Lineaire regressie en classificatie zijn de eenvoudigste vormen van gesuperviseerd leren, maar andere gesuperviseerde modellen, zoals “support vector machines” en “random forests”, worden ook vaak toegepast.

ONGESUPERVISEERD LEREN

Het doel van ongesuperviseerd leren (unsupervised learning) is dat het systeem zelf patronen in de data identificeert, terwijl gesuperviseerd leren een proces is van het in kaart brengen van verbanden tussen datapunten, zoals bij de vergelijking van twee beelden waarbij de objecten op de ene reeds zijn geïdentificeerd. Bij ongesuperviseerd leren worden patronen en structuren geïdentificeerd door het meten van de dichtheid of de overeenkomsten van datapunten in de dataset. Een veel voorkomende toepassing van ongesuperviseerd leren is clusteren. Daarbij ontvangt het model ongelabelde invoerdata en bepaalt het overeenkomsten en verschillen tussen de datapunten, wat resulteert in clusters op basis van gelijksoortige kenmerken, die belangrijke factoren zijn bij het categoriseren van de invoerdata. In het bovenstaande voorbeeld ontvangt het model verschillende soorten fruit, dieren, een bloem en een boom. Op basis van eigenschappen die uniek zijn voor elk van de categorieën, kan clustering dieren, fruit en planten scheiden in drie afzonderlijke clusters. Dimensionaliteitsreductie is een andere vorm van leren zonder toezicht.

BEKRACHTIGINGSLEREN



Bij bekrachtigingsleren (reinforcement learning) leren modellen eerder op basis van hun interacties met een virtuele of reële omgeving dan op basis van bestaande data. Bekrachtigingslerende "agenten" zoeken een optimale manier om een taak te volbrengen door een reeks stappen te nemen die de kans maximaliseren dat die taak volbracht wordt. Afhankelijk van het succes of falen van deze genomen stappen, worden zij beloond of bestraft. Deze "agenten" zijn geprogrammeerd om hun stappen zo te kiezen dat hun beloning maximaal is. Zij "leren" van beloningen en mislukkingen uit het verleden, verbeteren door vallen en opstaan, en kunnen langetermijnstrategieën ontwikkelen om hun beloning in het algemeen te maximaliseren in plaats van alleen naar hun volgende stap te kijken. Voorbeelden van bekrachtigingsleren zijn terug te vinden in de ontwikkeling van autonome voertuigen (zelfrijdende auto's). Bekrachtigingsleren wordt gebruikt om de prestaties van het voertuig in een gesimuleerde omgeving te verbeteren, door het te testen op zaken als de reactie op verkeerscontroles en versnelling. Door deze interacties met de gesimuleerde omgeving worden de bekrachtigingslerende “agenten” gestraft dan wel beloond op basis van de voltooiing van de taak, waardoor de toekomstige prestatie van het voertuig beïnvloed wordt.

FASES IN DE LEVENSCYCLUS VAN AI ONTWERP

AI ONTWERP

Project Planning

Een projectteam moet vanaf het begin beslissen wat de doelstellingen van het project zijn. Taken in deze fase kunnen bestaan uit het betrekken van belanghebbenden, uitvoeren van effectbeoordelingen, het in kaart brengen van belangrijke fases binnen het project, of een beoordeling van middelen en capaciteiten binnen het team of de organisatie. Een AI-projectteam beslist bijvoorbeeld of het al dan niet een AI-toepassing zal gebruiken in een agrarische setting om te voorspellen welke velden de volgende vijf jaar waarschijnlijk voor akkerbouw vatbaar zullen zijn, en wat de mogelijke gewasopbrengst zal zijn. Deze planning stelt het projectteam in staat na te denken over de ethische, sociaal-economische, juridische en technische aspecten, alvorens middelen te investeren in de ontwikkeling van het systeem.

Probleemstelling

Een projectteam moet bepalen welk probleem hun model zal behandelen, en welke invoerdata nodig zijn en voor welk doel. Het team moet de ethische en juridische implicaties van het gebruik van de data in overweging nemen en een grondige beschrijving geven van de bedoelde en onbedoelde gevolgen van het gebruik. Het team heeft bijvoorbeeld bepaald dat het overkoepelende thema van het project gewasopbrengsten zal zijn. Deze preciezere formulering helpt om een specifieke vraag te identificeren die aangekaart kan worden met behulp van data en kan verzekeren dat het resultaat in overeenstemming is met ethische en juridische overwegingen, zoals biodiversiteit of landgebruik.

Data-extractie of -verkrijging

Dit stadium omvat de processen van dataverzameling met het oog op het probleem in kwestie. Data-extractie kan via “web scraping” of dataregistratie via enquêtes of soortgelijke methoden, terwijl voor dataverkrijging een beroep kan worden gedaan op wettelijke overeenkomsten om reeds bestaande datasets te verkrijgen. In ons voorbeeld heeft het team besloten dat hun probleem het bepalen van factoren zal inhouden die van belang zijn voor het voorspellen van gewasopbrengsten in een bepaald landbouwseizoen. Zij besluiten data op te vragen bij een overheidsinstantie en bij landbouwcoöperaties, die beide wettelijke overeenkomsten vereisen voor het delen van data.

Data-Analyse

In dit stadium kan het projectteam beginnen met de inspectie van de data/gegevens. Dit zal in de eerste plaats een hoge mate van verkennende data-analyse (EDA - Exploratory Data Analysis) inhouden. EDA omvat het begrijpen van de samenstelling van de data door middel van visualisatie en samenvattende statistieken. Enkele vragen in deze fase kunnen zijn: zijn er ontbrekende data (onvolledige data), uitschieters (onverwachte data), onevenwichtige klassen (onevenwichtige data), of is er correlatie? Het team maakt bijvoorbeeld visualisaties om inzicht te krijgen in zaken zoals de verdeling van gewastypen over boerderijen, weersomstandigheden, pH-waarden van de bodem, naast het begrijpen van eventueel ontbrekende data.

Voorverwerking

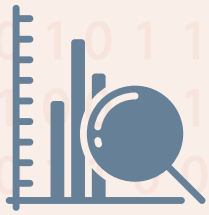
De voorverwerkingsfase is vaak het meest tijdrovende onderdeel van de ontwikkelingsfase van de AI-levenscyclus. Voorverwerking omvat taken zoals “data cleaning” (het herformatteren of verwijderen van onvolledige informatie) en “data wrangling” (het omzetten van data in een formaat dat bevorderlijk is voor het modelleren), naast andere processen die bijdragen aan het trainingsproces van het model. Tijdens de voorverwerking merken de leden van het team bijvoorbeeld op dat de pH-niveaus van de bodem zowel als numerieke en als tekststring data behandeld worden , wat problemen zou veroorzaken bij het uitvoeren van het model. Zij besluiten daarom alle pH-niveaus van de bodem hetzelfde datatype te maken door de tekststring data om te zetten naar numerieke data.

Modelselectie en -Training

Modellen moeten specifiek geselecteerd worden voor het probleem dat in de ontwerpfase werd vastgesteld. Model-types variëren in complexiteit; bij de modelselectie worden echter ook andere factoren in aanmerking genomen, zoals het type data, de hoeveelheid en de beschikbaarheid ervan. Modellen die niet voldoende complex zijn, lopen het risico op “underfitting” (= zij houden onvoldoende rekening met de data). Voorverwerkte dataworden gesplitst in trainings- en testsets om “overfitting” te voorkomen. “Overfitting” doet zich voor wanneer het model de trainingsdata te nauwkeurig weerspiegelt en niet in staat is nieuwe, "ongeziene" data in te passen om nauwkeurige voorspellingen te doen voor inputdata die niet in de trainingsset zaten. Trainingsdata worden gebruikt om de parameters van het geselecteerde model te verfijnen. Als voorbeeld van modelselectie heeft het projectteam besloten een lineair regressiemodel te gebruiken om aan de hand van data uit het verleden toekomstige gewasopbrengsten te voorspellen. Men wilde een model dat interpreteerbaar was om de resultaten volledig te kunnen verklaren, daarom was de keuze voor een eenvoudige techniek als lineaire regressie logisch.

Testen en Valideren van het Model

Na de training wordt het model afgesteld en getest met "ongeziene" data. Validatiesets worden gebruikt om aspecten van het model op hoger niveau aan te passen (zoals hyperparameters die bepalen hoe het model leert) en worden vaak gecreëerd door de dataset aanvankelijk in drie delen op te splitsen, bijvoorbeeld 60% trainingsdata, 20% testdata en 20% validatiedata. Tijdens de validatie kunnen elementen van de architectuur van het model worden gewijzigd om de prestaties van het model te beïnvloeden. In ons voorbeeld realiseert het team bij het uitvoeren van het model dat het aantal variabelen in het model tot “overfitting” leidt. Het team besluit dan een regularisatieterm (een methode om de fout van het model te verminderen) toe te voegen om onbelangrijke variabelen te verwijderen. Het model wordt vervolgens getest op onbekende data om de toepassing in de echte wereld na te bootsen en de prestaties en nauwkeurigheid te bevestigen.



Modelrapportering

Nadat het team het model heeft getraind, gevalideerd en getest, moet de evaluatie van het model (met inbegrip van diverse prestatie- en effectmetingen), samen met gedetailleerde informatie over de workflow van het model, worden opgesteld om transparante discussies over de output van het model beter te ondersteunen. Om de ontwikkelingsfase af te ronden, legt het team bijvoorbeeld diverse prestatiemetingen van hun model vast, samen met de processen om tot de huidige iteratie van het model te komen, met inbegrip van de voorverwerking en de beslissing om in de fase van het testen en valideren van het model regularisatie toe te voegen.

GEBRUIK

Implementatie van het Model

De volgende fase van de AI-levenscyclus omvat de inzet van het getrainde model in de echte wereld. Door een doeltreffende implementatie kan het model in een groter systeem worden opgenomen. Nieuwe data worden door het geïmplementeerde model verwerkt om het beoogde doel te bereiken, dat in de ontwerpfase is bepaald. Ons AI-projectteam uit het voorbeeld heeft besloten dat hun model voor gewasopbrengst klaar is voor gebruik. Ze kiezen ervoor om het beschikbaar te maken voor verschillende landbouwcoöperaties en vragen hen om het op hun data toe te passen om te zien of het bruikbare inzichten oplevert.

Opleiding van gebruikers

De gebruikers van het systeem moeten worden opgeleid om de logica van het systeem te begrijpen, om de beslissingen ervan in gewone taal te kunnen uitleggen aan degenen voor wie de beslissingen zijn bestemd, en om een onafhankelijk en onbevooroordeeld oordeel te kunnen vellen over de kwaliteit, betrouwbaarheid en billijkheid van de outputdata van het systeem. Nadat het team bijvoorbeeld specifieke gebruikers in de landbouwsector heeft opgeleid in het gebruik van hun model, zullen deze gebruikers verslag uitbrengen over de vraag of zij het systeem nuttig, betrouwbaar en accuraat vinden, naast andere meetcriteria.

Toezicht

Nadat het model door het team is geïmplementeerd, moet het worden gecontroleerd om ervoor te zorgen dat het nog steeds het gewenste doel dient, dat het op verantwoorde wijze en binnen het beoogde toepassingsgebied wordt gebruikt, en dat het reageert op nieuwe omstandigheden in de praktijk. Het team merkt bijvoorbeeld dat een nieuwe variabele om de waterkwaliteit te meten door een normalisatie-instelling is vrijgegeven. Dit zou kunnen leiden tot een gebrek aan standaardisering in de data, aangezien het geen oorspronkelijke variabele was die deel uitmaakte van de reeks trainingsdata. Zij besluiten deze wijziging in het model op te nemen om in overeenstemming te blijven met de landbouwnormen en -praktijken.

Bijwerken of Afvoeren

Na verloop van tijd kan het model aan doeltreffendheid inboeten, zodat het toezichthoudende team eerdere stadia van de ontwikkelingsfase, waaronder modelselectie en -opleiding, opnieuw moet bekijken. Als er meer ingrijpende wijzigingen nodig zijn, moet het systeem mogelijk worden afgevoerd, en dient het ontwerpproces met de projectplanning opnieuw opgestart te worden. Het team uit ons voorbeeld heeft het model verscheidene malen moeten hertrainen op basis van nieuwe variabelen en niet-gestandaardiseerde datasets. Zij blijven het model opvolgen terwijl zij alternatieve opties overwegen, waaronder de ontwikkeling van een nieuw systeem.

03 EEN KORTE INLEIDING TOT MENSENRECHTEN, DEMOCRATIE EN DE RECHTSSTAAT

"Alle mensenrechten zijn universeel, ondeelbaar, onderling afhankelijk en onderling verbonden"

-Verklaring van Wenen van de Verenigde Naties, 1993

Mensenrechten, democratie en de rechtsstaat zijn nauw met elkaar verbonden. Het vermogen van legitieme regeringen om de mensenrechten daadwerkelijk te beschermen is gebaseerd op de onderlinge afhankelijkheid van robuuste en verantwoordingsplichtige democratische instellingen, inclusieve en transparante besluitvormingsmechanismen, en een onafhankelijke en onpartijdige rechterlijke macht die de rechtsstaat waarborgt. In het algemeen zijn mensenrechten de elementaire rechten en vrijheden die iedere persoon ter wereld van de wieg tot het graf bezit en die de onschendbare waardigheid van ieder individu, ongeacht ras, etniciteit, geslacht, leeftijd, seksuele geaardheid, klasse, godsdienst, handicap, taal, nationaliteit of enig ander toegeschreven kenmerk, garanderen en beschermen. Deze fundamentele rechten en vrijheden scheppen verplichtingen die regeringen ertoe verbinden de mensenrechten te eerbiedigen, te beschermen en te realiseren. Bij niet-nakoming van deze verplichtingen beschikken individuen over rechtsmiddelen die het mogelijk maken eventuele schendingen van de mensenrechten te herstellen.

MENSENRECHTEN IN EEN OOGOPSLAG

Historisch gezien zijn de basisrechten en -beginselen die bekend zijn geworden als mensenrechten, voor het eerst ontstaan in het midden van de 20ste eeuw, in de nasleep van de wreedheden en het trauma van de Tweede Wereldoorlog.

1948

De Verenigde Naties nemen de **Universele Verklaring van de Rechten van de Mens** (UVRM) aan, die een eerste internationale norm voor fundamentele rechten en vrijheden biedt. Hoewel niet juridisch bindend, ligt dit document aan de basis van de talrijke verdragen, conventies en handvesten met betrekking tot mensenrechten die tot op heden wereldwijd zijn aangenomen.

1953

Het Europees Verdrag tot bescherming van de Rechten van de Mens (EVRM) treedt in werking. Dit internationale verdrag, dat voor het eerst werd opgesteld door de Raad van Europa in 1950, omvat de burgerlijke en politieke rechten waaraan de 47 lidstaten van de Raad wettelijk gebonden zijn. Naast het vastleggen van basisrechten die dienen om de onschendbare waardigheid van elke persoon te vrijwaren, legt het EVRM regeringen ook verplichtingen op om gewone mensen te beschermen tegen schendingen van hun mensenrechten.

1961

De Raad van Europa stelt zijn **Europees Sociaal Handvest** (ESH) open voor ondertekening. Dit verdrag breidt de basisrechten uit met sociale en economische rechten die gezondheid, arbeidsomstandigheden, huisvesting, migrantenarbeid, gelijkheid van mannen en vrouwen, en sociale zekerheid omvatten. In 1988 werden aanvullende protocollen toegevoegd ter versterking van gelijke kansen op de werkvloer, werknemersparticipatie en bescherming van armen en ouderen. In 1996 werd een herzien ESH aangenomen.

1966

De VN neemt het **Internationaal Verdrag inzake burgerrechten en politieke rechten** (IVBPR/BUPO-verdrag) en het **Internationaal Verdrag inzake economische, sociale en culturele rechten** (IVESCR/ECOSOC-verdrag) aan. Het IVBPR omvat het verbod op foltering, het recht op een eerlijk proces, non-discriminatie en privacyrechten. Het IVESCR breidt de basisrechten uit met het recht op rechtvaardige arbeidsomstandigheden, gezondheid, levensstandaard, onderwijs en sociale zekerheid. De UVRM, het IVBPR en het IVESCR van de VN staan nu samen bekend als de **Internationale "Bill of Human Rights".**

2009

Het Handvest van de grondrechten van de Europese Unie wordt volledig van kracht door het Verdrag van Lissabon. Dit codificeert een basispakket van burgerlijke, politieke, sociale, economische en culturele rechten voor burgers van de Europese Unie in het EU-recht. Tot de mensenrechten die onder het Handvest vallen, behoren diegene die betrekking hebben op menselijke waardigheid, fundamentele vrijheden, gelijkheid, solidariteit en economische rechten, en het recht op deelname aan het gemeenschapsleven.

TWEE FAMILIES MENSENRECHTEN

Het geheel van beginselen dat de mensenrechten vormt, kan in twee groepen worden opgesplitst:

Burgerlijke en Politieke Rechten

Belangrijkste rechten:

-Recht op leven en menselijke waardigheid
-Recht op lichamelijke en geestelijke integriteit
-Recht op vrijheid en veiligheid van personen
-Vrijwaring van foltering en wrede behandeling
-Recht op een eerlijk proces en een eerlijke rechtsgang
-Recht op een daadwerkelijk rechtsmiddel
-Vrijheid van gedachten, geweten en godsdienst
-Vrijheid van meningsuiting en van mening
-Recht op eerbiediging van het privé-leven en het familie- en gezinsleven
-Recht op bescherming van persoonsgegevens
-Recht op non-discriminatie
-Recht op gelijkheid voor de wet
-Vrijheid van vergadering en vereniging
-Recht om deel te nemen aan het openbaar bestuur

Sociale, Economische en Culturele rechten

Belangrijkste rechten:

-Recht op rechtvaardige, veilige en gezonde arbeidsomstandigheden
- Recht op eerlijke beloning
- Recht op beroepsopleiding
- Recht op gelijke kansen op de werkplek
- Recht op vereniging en collectieve onderhandeling
- Recht op sociale zekerheid
- Recht op onderwijs
- Recht op een adequate levensstandaard
- Recht op sociale en medische bijstand
- Recht op bescherming van de gezondheid
- Recht op bescherming voor arbeidsmigranten
- Recht op sociale bescherming voor ouderen
- Recht op bescherming tegen sexuele intimidatie
- Recht op bescherming tegen armoede en sociale uitsluiting

DE ONDERLINGE AFHANKELIJKHEID VAN MENSENRECHTEN, DEMOCRATIE EN DE RECHTSSTAAT

De onderlinge afhankelijkheid van mensenrechten, democratie en de rechtsstaat vindt haar oorsprong in hun verweven en symbiotische aard. De legitimiteit van democratische instellingen is geworteld in de gedachte dat iedere burger in gelijke mate het recht heeft deel te nemen aan het gedeelde gemeenschapsleven en aan de sturing van de collectieve besluiten die daarop van invloed zijn. Willen burgers dit recht op deelneming aan het openbare leven kunnen uitoefenen, dan moeten zij eerst vele andere onderling samenhangende burgerlijke, politieke, sociale, culturele en economische rechten bezitten:

- Zij moeten vrijheid van gedachte, vereniging, vergadering en meningsuiting genieten.
- Zij moeten voor de wet op voet van gelijkheid worden behandeld en beschermd worden tegen elke vorm van discriminatie die hun volledige en billijke deelneming aan het gemeenschapsleven zou kunnen belemmeren.
- Zij moeten toegang hebben tot de materiële middelen voor participatie door de voorziening van behoorlijk onderwijs, passende levens- en arbeidsomstandigheden, gezondheid, veiligheid en sociale zekerheid.
- Zij moeten toegang hebben tot doeltreffende rechtsmiddelen in geval een van hun fundamentele rechten wordt geschonden.

In dit laatste opzicht vormt de rechtsstaat de institutionele basis voor het waarborgen van zowel democratische participatie als de bescherming van fundamentele rechten en vrijheden. Een onafhankelijke en onpartijdige rechterlijke macht, die de burgers een behoorlijke rechtsgang en een eerlijke en gelijke behandeling door de wet waarborgt, vormt een waarborg voor verhaal wanneer fundamentele rechten of vrijheden geschonden zouden kunnen worden.

04 KANSEN EN RISICO'S VAN AI/ML EN HUN IMPACT OP MENSENRECHTEN, DEMOCRATIE EN DE RECHTSSTAAT

Technologieën op het gebied van artificiële intelligentie (AI) bieden tal van mogelijkheden voor de verbetering van het leven van de mens en het functioneren van de overheid. De kracht, schaal en snelheid van AI-systemen kunnen de efficiëntie en doeltreffendheid op tal van gebieden, zoals gezondheidszorg, vervoer, onderwijs en openbaar bestuur verbeteren. Ze kunnen vervelende, gevaarlijke, onaangename en ingewikkelde taken overnemen van menselijke werknemers. AI-technologieën kunnen echter ook een negatieve impact hebben op mensenrechten, de democratie en de rechtsstaat. Deze gecombineerde mogelijkheden en risico's moeten worden begrepen in het licht van het feit dat AI **"sociaal-technisch"** van aard is - AI is een brede waaier van gesofisticeerde technologieën die opereren in een menselijke context, ontworpen om door mensen bepaalde doelstellingen te verwezenlijken. Als zodanig kan worden gesteld dat AI-technologieën de waarden en keuzes van de mensen die ze bouwen en gebruiken weerspiegelen.

AI kan worden gebruikt om voorspellingen te doen over menselijk gedrag, om ziekte-indicatoren te identificeren en om risico's voor de belangen of het welzijn van anderen te beoordelen. Al deze taken kunnen een invloed hebben op de rechten, de kansen en het welzijn van degenen op wie ze worden toegepast. Toerekenbaarheid is dan ook een essentieel aspect van de ontwikkeling en het gebruik van dergelijke systemen. Hoewel AI vervelende of complexe taken van de mens kan overnemen, kunnen de keuzes die bij de constructie en het gebruik van AI-systemen worden gemaakt, resulteren in de reproductie van schadelijke vooroordelen (of “bias”) en andere menselijke beoordelingsfouten die nadelige gevolgen hebben voor de betrokken personen en de ruimere samenleving op manieren die moeilijker te identificeren zijn dan wanneer zij door mensen zouden worden uitgevoerd.

Behalve de evaluatie van de technische kenmerken van een bepaald systeem of een bepaalde technologie, vereist verantwoording ten aanzien van AI dan ook dat grondig dient te worden nagedacht over de mogelijke nadelen en voordelen voor individuen en groepen. Een van de potentiële nadelen is onrechtvaardige vooringenomenheid, die expliciet kan voorkomen, bijvoorbeeld wanneer AI-modellen discriminerende voorspellingen doen of een bepaalde demografische groep of identiteit anders behandelen dan anderen zonder enige rechtvaardiging. AI-systemen beoordelen op hun potentieel om schade te berokkenen, wordt bemoeilijkt door de ondoorzichtigheid van sommige AI-systemen. Het werk van AI-technologieën wordt niet alleen ontwikkeld met behulp van gespecialiseerde kennis, maar kan ook moeilijk te interpreteren of uit te leggen zijn zowel omwille van de technische complexiteit als van de bescherming van intellectuele eigendom.

De implicaties van AI-systemen voor de mensenrechten kunnen worden gezien in het licht van de bepalingen van het Europees Verdrag tot Bescherming van de Rechten van de Mens (EVRM) en het Europees Sociaal Handvest (ESH), met inbegrip van de daarin vervatte specifieke waarborgen inzake **vrijheid en rechtvaardigheid, privacy, vrijheid van meningsuiting, gelijkheid en non-discriminatie, en sociale en economische rechten.** AI heeft nog andere gevolgen voor de democratie en de rechtsstaat die niet duidelijk onder de bepalingen van het EVRM en het ESH vallen, maar die desondanks even belangrijk zijn. Een grondige afweging van de risico's en kansen van AI-systemen zal ons helpen te bepalen waar bestaande rechten en vrijheden de nodige bescherming bieden, waar verdere verduidelijking van bestaande rechten en vrijheden nodig is, en waar nieuwe rechten en vrijheden moeten worden afgestemd op de nieuwe uitdagingen en kansen die AI en machinaal leren met zich meebrengen.

Vrijheid en rechtvaardigheid: AI kan een negatieve impact hebben op de vrijheid en rechten van individuen, met name wanneer het wordt geïmplementeerd in contexten met een grote impact, zoals het strafrecht. De complexiteit en de ondoorzichtigheid van AI-systemen kunnen een belemmering vormen voor het recht op een eerlijk proces, met inbegrip van het recht op procedurele gelijkheid, waarbij een partij die het voorwerp is van een algoritmisch besluit, de redenering naar behoren kan onderzoeken en betwisten. Hoewel het gebruik van AI in deze context willekeur en discriminerend optreden kan verminderen, kunnen rechterlijke beslissingen die door AI worden ondersteund of geïnformeerd, een negatieve invloed hebben op de jurisprudentie en de onafhankelijkheid van de rechterlijke macht. Daarom moeten de justitiële actoren voldoende inzicht hebben in de AI die zij gebruiken om ervoor te zorgen dat zij verantwoording af kunnen leggen voor beslissingen die met behulp van AI zijn genomen.

Een systeem dat strafrechtelijke beslissingen ondersteunt met scores die het risico weergeven dat een veroordeelde nieuwe misdrijven zal plegen, moet interpreteerbaar en controleerbaar zijn en moet door de verdachte kunnen worden aangevochten om een eerlijk en open proces te waarborgen.

Een systeem dat gezichtsuitdrukkingen, toon van de stem, woordkeuze en andere biometrische kenmerken analyseert en vergelijkt met modellen om te voorspellen of een sollicitant "succesvol" zal zijn, kan het gevoel van lichamelijke en emotionele privacy van de sollicitant schenden.

Privacy: AI heeft toegang tot enorme hoeveelheden gegevens over personen en kan deze met ongelofelijke snelheid verwerken. AI kan voorspellingen doen over iemands gedrag, gemoedstoestand en identiteit door informatie op te sporen die niet noodzakelijkerwijs als persoonlijk of privé wordt beschouwd, zoals gezichtsuitdrukkingen, hartslag, fysieke locatie en andere schijnbaar alledaagse of openbaar toegankelijke data. Dit kan een inbreuk vormen op iemands gevoel van privacy, en kan ook zogenaamde "panoptische effecten" hebben doordat mensen hun gedrag veranderen als zij vermoeden dat zij worden geobserveerd of geanalyseerd.

Vrijheid van meningsuiting, vereniging en vergadering: Een goed functionerende democratie vereist een open sociaal en politiek discours en de minimalisering van ongepaste beïnvloeding of manipulatie door een bepaalde persoon of institutie. AI brengt deze waarden in gevaar wanneer het wordt gebruikt voor het verzamelen en verwerken van informatie over online- en offline activiteiten door het bijhouden en analyseren van website- en social mediagebruik of het extraheren van informatie via biometrische surveillance. AI die op deze manier wordt gebruikt, draagt bij tot het gevoel dat iemand in de gaten wordt gehouden en wordt afgeluisterd, waardoor vrije meningsuiting en politieke actie mogelijk worden belemmerd. AI die door sociale media platformen wordt gebruikt, bepaalt welke berichten en advertenties worden getoond, gericht op individuele interesses en vooroordelen om mensen op het platform te houden, terwijl potentieel verdeeldheid zaaiende, antidemocratische of gewelddadige wereldbeelden worden versterkt. AI wordt ook ingezet om zeer realistische maar fake video's, nep accounts en andere gefabriceerde inhoud te produceren die iemands vermogen om geïnformeerde meningen te vormen, kunnen belemmeren.

Systemen voor live gezichtsherkenning kunnen burgers beperken in hun vrijheid van vergadering, hen beroven van de bescherming van anonimiteit en een 'chilling' effect hebben op sociale solidariteit en democratische participatie. Op AI gebaseerd biometrisch toezicht kan burgers ook het recht op geïnformeerde en uitdrukkelijke toestemming voor het verzamelen van persoonsgegevens ontnemen.

Gelijkheid en non-discriminatie: AI-systemen zijn in staat patronen van discriminatoir handelen die bestaan in de samenleving waarin zij worden gecreëerd en gebruikt, te reproduceren en te vergroten. Dit kan gebeuren wanneer de stereotiepe vooroordelen en blinde vlekken van systeemontwikkelaars de keuzes bepalen bij het ontwerpen en inzetten van de systemen. Dit kan ook gebeuren wanneer historische structuren van ongelijkheid en discriminatie verankerd raken in de datasets die worden gebruikt om AI- en machine learning modellen te trainen. Wanneer AI vertrouwt op dergelijke bevooroordeelde informatie, kunnen discriminerende menselijke beslissingen die een dataset hebben voortgebracht, leiden tot discriminerende algoritmische beslissingen en gedragingen.

Wanneer predictive policing systemen uitgaan van historische data, bestaat het risico dat zij de resultaten van eerdere discriminerende praktijken reproduceren. Dit kan leiden tot "feedback loops", waarbij elk nieuw besluit op basis van historische data nieuwe data oplevert, waardoor leden van gemarginaliseerde groepen buitenproportioneel vaak worden verdacht en gearresteerd.

"Ride-hailing" en bezorgingsdiensten, gecoördineerd door mobiele apps, maken het mogelijk voor bedrijven om het management van en het toezicht op grote groepen werkenden te automatiseren en de arbeidsverhoudingen en managementpraktijken op hun beurt te dehumaniseren. Dit kan de positie van werknemers aantasten en hun verhaalsmogelijkheden beperken bij onjuiste algoritmische managementbeslissingen over het loon of het werk.

Sociale en economische rechten: AI-systemen worden door werkgevers en overheden steeds vaker gebruikt op manieren die sociale en economische rechten in gevaar brengen. Werkgevers gebruiken technologie om het gedrag van werknemers te controleren, vakbondsvorming te verstoren en beslissingen te nemen over werving, loon en promotie. In sommige werkomgevingen worden mensen voornamelijk gemanaged door algoritmische besluitvormingssystemen, waardoor hun economische kansen in het gedrang kunnen komen. Ook de invloed van de overheid op economische welvaart wordt beïnvloed wanneer AI wordt gebruikt om overheidsuitkeringen en gezondheidszorg toe te wijzen. Onvoldoende toezicht op een dergelijk proces kan leiden tot weigering van uitkeringen aan degenen die er recht op hebben, waardoor hun welzijn wordt bedreigd. De automatisering van zowel de voorwaarden als de toekenning van een uitkering, kan tot een efficiëntere dienstverlening leiden. Het kan er ook toe leiden dat geweigerden geen verhaalsmogelijkheden hebben of dat zij zich zonder hulp een weg moeten banen door complexe formulieren en andere processen. Naast deze zorgen over mensenrechten is er het probleem van machtsconcentratie die AI verschaft aan zijn meest invloedrijke ontwikkelaars en uitvoerders in de private en de publieke sector. De exploitanten van grote online platformen gebruiken AI om te beslissen welke inhoud wordt getoond en wie de luidste stem krijgt ten behoeve van hun eigen belangen, in plaats het belang van de democratie. Overheden gebruiken AI om informatie te rangschikken en te ordenen en om burgers te controleren en te tracken. Of het nu gaat om bedrijven of overheden, AI kan gebruikt worden om meningen te vormen en afwijkende meningen te onderdrukken.

In antwoord op deze overwegingen en bekommernissen moeten overheden het voorzorgsbeginsel toepassen bij regulering van AI, zodat de mogelijkheden die AI biedt kunnen worden verwezenlijkt terwijl de risico's voor mensen en menselijke belangen zoveel mogelijk worden beperkt. In contexten waarin het voorzorgsbeginsel ontoereikend is om de risico's te beperken, dienen overheden een verbod op het gebruik van AI te overwegen. Wanneer er onzekerheid bestaat over het niveau of de gevolgen van potentiële risico's, moeten regeringen meer regelgevend toezicht en controle op AI-systemen uitoefenen en bereid zijn het gebruik ervan te verbieden.

05 BEGINSELEN EN PRIORITEITEN VOOR EEN JURIDISCH KADER

In september 2019 heeft het Comité van Ministers van de Raad van Europa de opdracht voor het Ad Hoc Comité voor Artificiële Intelligentie (CAHAI) vastgesteld. CAHAI is belast met het onderzoeken van de haalbaarheid en de mogelijke elementen van een juridisch kader voor de ontwikkeling, het ontwerp en het gebruik van AI-systemen, op basis van normen van de Raad van Europa gebieden van mensenrechten, democratie en de rechtsstaat. Als eerste en noodzakelijke stap in de uitvoering van deze opdracht heeft CAHAI in zijn *Haalbaarheidsstudie*, in december 2020 in plenaire vergadering aangenomen, negen beginselen en prioriteiten voorgesteld die ten grondslag moeten liggen aan een dergelijk kader van bindende en niet-bindende juridische instrumenten:

MENSELIJKE WAARDIGHEID

Alle individuen zijn inherent en onschendbaar waardig op grond van hun hoedanigheid als mens. Mensen moeten worden behandeld als morele subjecten, en niet als objecten die algoritmisch te beoordelen of te manipuleren zijn.

MENSELIJKE VRIJHEID & AUTONOMIE

Mensen moeten in staat worden gesteld om op een geïnformeerde en autonome manier te bepalen of, wanneer en hoe AI-systemen worden gebruikt. Deze systemen mogen niet worden gebruikt om mensen te conditioneren of te controleren, maar moeten juist hun vermogens verrijken.

VOORKOMEN VAN SCHADE

De fysieke en mentale integriteit van de mens en de duurzaamheid van de biosfeer moeten worden beschermd, en er moeten aanvullende voorzorgsmaatregelen worden genomen om kwetsbaren te beschermen. AI-systemen mogen geen negatieve invloed hebben op het menselijk welzijn of de gezondheid van de planeet.

NON-DISCRIMINATIE, GENDERGELIJKHEID, BILLIJKHEID & DIVERSITEIT

Alle mensen hebben het recht op non-discriminatie en het recht op gelijkheid en gelijke behandeling voor de wet. AI-systemen moeten zo worden ontworpen dat zij eerlijk, billijk en inclusief zijn in de positieve effecten en in de spreiding van de risico's.

TRANSPARANTIE EN UITLEGBAARHEID VAN AI-SYSTEMEN

Wanneer een product of dienst gebruik maakt van een AI-systeem, moet hierover open worden gecommuniceerd aan de betrokken personen. Ook moet zinvolle informatie worden verstrekt over de achterliggende redenen van de output van het AI-systeem.

GEGEVENSBESCHERMING EN RECHT OP PRIVACY

Het ontwerp en het gebruik van AI-systemen die gebaseerd zijn op de verwerking van persoonsgegevens, moeten het recht op het privé- en gezinsleven waarborgen, met inbegrip van het recht van het individu om zijn eigen data te beheren. Geïnformeerde, vrijwillige en ondubbelzinnige toestemming moet daarbij een rol spelen.

VERANTWOORDING EN VERANTWOORDELIJKHEID

Alle personen die betrokken zijn bij het ontwerp en de inzet van AI-systemen moeten ter verantwoording worden geroepen wanneer toepasselijke rechtsnormen worden geschonden of wanneer eindgebruikers of anderen ongerechtvaardigde schade wordt berokkend. Diegenen die de negatieve gevolgen ondervinden, moeten een effectieve mogelijkheid tot verhaal hebben.

DEMOCRATIE

Transparante en inclusieve toezichtmechanismen moeten ervoor zorgen dat de democratische besluitvormingsprocessen, het pluralisme, de toegang tot informatie, de autonomie en de economische en sociale rechten worden beschermd in de context van het ontwerp en het gebruik van AI-systemen.

RECHTSSTAAT

AI-systemen mogen een eerlijk proces en de onafhankelijkheid en onpartijdigheid van de rechterlijke macht niet ondermijnen. Daartoe moeten de transparantie, de integriteit en de billijkheid van data en verwerkingsmethoden worden gewaarborgd.

BEGINSELEN, RECHTEN, EN PLICHTEN MET ELKAAR VERBINDEN

Deze negen beginselen en prioriteiten zijn **horizontaal toepasbaar**. Zij zijn van toepassing op het ontwerp, de ontwikkeling en de inzet van AI-systemen in **alle sectoren en gebruikssituaties**, maar kunnen worden gecombineerd met een sectorspecifieke aanpak die (meer gedetailleerde) contextuele eisen bevat in de vorm van "soft law", zoals sectorale standaarden, richtsnoeren of beoordelingslijsten.

Het is de bedoeling dat het juridisch kader vertrekt vanuit deze brede invalshoek. Het zal erop gericht zijn de negen beginselen en prioriteiten te waarborgen door **concrete rechten** te identificeren die de verwezenlijking van deze sectoroverschrijdende beginselen op individueel niveau waarborgen en de **belangrijkste verplichtingen en eisen** waaraan ontwikkelaars en gebruikers moeten voldoen bij de ontwikkeling en het gebruik van AI-systemen in overeenstemming met mensenrechten, democratie en de rechtsstaat. De geïdentificeerde rechten kunnen (1) rechtstreeks volgen uit bestaande rechten, (2) nieuw gecreëerde rechten zijn, afgestemd op de uitdagingen en kansen van AI, of (3) een verdere verduidelijkingen van bestaande rechten behelzen.

Hier wordt in kaart gebracht hoe elk van de beginselen en prioriteiten is verbonden met de overeenkomstige rechten en verplichtingen:

	MATERIËLE RECHTEN	BELANGRIJKSTE VERPLICHTINGEN
MENSELIJKE WAARDIGHEID	- Het recht op menselijke waardigheid, het recht op leven **(Art. 2 EVRM)**, en het recht op lichamelijke en geestelijke integriteit. - Het recht geïnformeerd te worden over het feit dat men met een AI-systeem in plaats van met een mens te maken heeft. - Het recht om interactie met een AI-systeem te weigeren wanneer dit de menselijke waardigheid zou kunnen aantasten.	- Lidstaten moeten ervoor zorgen dat taken die de menselijke waardigheid dreigen te schenden indien ze door machines in plaats van door mensen worden uitgevoerd, aan mensen worden voorbehouden. - Lidstaten moeten van de aanbieders van AI-systemen eisen dat zij mensen op de hoogte brengen als zij met een AI-systeem en niet met een mens te maken hebben in elke context waarin verwarring zou kunnen ontstaan.
MENSELIJKE VRIJHEID EN AUTONOMIE	- Het recht op vrijheid en veiligheid **(Art. 5 EVRM).** - Het recht op menselijke autonomie en zelfbeschikking. Het recht om niet onderworpen te worden aan een beslissing die uitsluitend op geautomatiseerde verwerking gebaseerd is wanneer dit juridische gevolgen teweegbrengt of soortgelijke significante gevolgen heeft voor natuurlijke personen. - Het recht om door een AI-systeem geïnformeerde en/of genomen besluiten daadwerkelijk te betwisten en aan te vechten en te eisen dat het besluit door een persoon getoetst wordt. - Een vrijwillig 'opt-out' recht van AI-gedreven manipulatie, geïndividualiseerde profilering en voorspelling. Dit geldt ook in geval van verwerking van niet-persoonsgegevens. - Het recht om, wanneer legitieme dwingende redenen dit niet verhinderen, de mogelijkheid te hebben om te kiezen voor contact met een mens in plaats van met een robot.	- Alle AI-gedreven manipulatie, geïndividualiseerde profilering en voorspellingen waarbij persoonsgegevens worden verwerkt, moeten voldoen aan de verplichtingen van het Verdrag van de Raad van Europa tot bescherming van personen met betrekking tot de geautomatiseerde verwerking van persoonsgegevens. - De lidstaten moeten de gemoderniseerde versie van het Verdrag ("Verdrag 108+") daadwerkelijk implementeren om problemen in verband met AI beter aan te kunnen pakken. - De lidstaten moeten de ontwikkelaars en aanbieders van AI verplichten mechanismen voor menselijk toezicht in te stellen die de menselijke autonomie beschermen, op een wijze die afgestemd is op de specifieke risico's die voortvloeien uit de context waarin het AI-systeem wordt ontwikkeld en gebruikt. - Lidstaten moeten AI-ontwikkelaars en -aanbieders ertoe verplichten de verhaalsmogelijkheden tijdig en naar behoren mee te delen.

	MATERIËLE RECHTEN	BELANGRIJKSTE VERPLICHTINGEN
VOORKOMEN VAN SCHADE	- Het recht op leven **(Art. 2 EVRM)** en het recht op lichamelijke en geestelijke integriteit. - Het recht op de bescherming van het milieu. - Het recht op duurzaamheid van de gemeenschap en de biosfeer.	- Lidstaten moeten ervoor zorgen dat ontwikkelaars en aanbieders van AI-systemen passende maatregelen nemen om fysieke of mentale schade aan personen, de samenleving en het milieu te minimaliseren. - Lidstaten moeten ervoor zorgen dat er adequate (door het ontwerp ingegeven) veiligheids-, beveiligings- en robuustheidsvereisten bestaan en dat de ontwikkelaars en gebruikers van AI-systemen zich daaraan houden. - Lidstaten moeten ervoor zorgen dat AI-systemen op duurzame wijze worden ontwikkeld en gebruikt, met volledige inachtneming van de toepasselijke normen voor milieubescherming.
NON-DISCRIMINATIE, GENDERGELIJKHEID, BILLIJKHEID EN DIVERSITEIT	- Het recht op non-discriminatie **(op basis van de beschermde gronden als bedoeld in artikel 14 van het EVRM en Protocol 12 bij het EVRM)**, met inbegrip van intersectionele discriminatie. - Het recht op non-discriminatie en het recht op gelijke behandeling. - AI-systemen kunnen ook aanleiding geven tot onrechtvaardige categorisering op basis van nieuwe vormen van differentiatie die traditioneel niet beschermd zijn. - Dit recht moet worden gewaarborgd met betrekking tot de gehele levenscyclus van een AI-systeem (ontwerp, ontwikkeling, implementatie en gebruik), alsmede met betrekking tot de menselijke keuzes betreffende AI-ontwerp, -inzet en -gebruik, ongeacht of dit in de publieke of in de private sector plaatsvindt.	- Lidstaten zijn verplicht ervoor te zorgen dat de AI-systemen die zij inzetten niet leiden tot onwettige discriminatie, schadelijke stereotypering (met inbegrip van, maar niet beperkt tot, genderstereotypering) en bredere sociale ongelijkheid, en moeten derhalve het hoogste toetsingsniveau toepassen wanneer zij AI-systemen gebruiken of het gebruik ervan promoten op gevoelige beleidsterreinen van de overheid, met inbegrip van, maar niet beperkt tot, rechtshandhaving, justitie, asiel en migratie, gezondheid, sociale zekerheid en werkgelegenheid. - Lidstaten moeten in de (aanbestedings)procedures voor overheidsopdrachten voor AI-systemen vereisten opnemen inzake non-discriminatie en bevordering van gelijkheid, en verzekeren dat de systemen onafhankelijk worden gecontroleerd op discriminatoire effecten voordat zij worden ingezet. - De lidstaten moeten vereisten vaststellen om de potentiële discriminatoire effecten van zowel door de overheid als door de private sector gebruikte AI-systemen doeltreffend tegen te gaan en om individuen te beschermen tegen de negatieve gevolgen daarvan. Deze vereisten moeten in verhouding staan tot de risico's. - Lidstaten moeten diversiteit en gendergelijkheid in het AI werkveld stimuleren, alsmede het ontvangen van periodieke feedback van een divers gamma van belanghebbenden. Bewustmaking van het risico van discriminatie, met inbegrip van nieuwe vormen van differentiatie, en van vooringenomenheid in de context van AI moet worden bevorderd.

	MATERIËLE RECHTEN	BELANGRIJKSTE VERPLICHTINGEN
TRANSPARANTIE EN UITLEGBAARHEID VAN AI-SYSTEMEN	- Het recht om onmiddellijkte worden geïnformeerd wanneer een beslissing met rechtsgevolgen of soortgelijke ingrijpende invloed op een individu, steunt op of wordt genomen door een AI-systeem **(Verdrag 108+).** - Het recht op een zinvolle uitleg over hoe een dergelijk AI-systeem functioneert, welke optimaliseringslogica het volgt, welk soort data het gebruikt, en hoe het belangen beïnvloedt, indien sprake is van rechtsgevolgen of beïnvloeding van de persoonlijke levenssfeer. De uitleg moet afgestemd zijn op de context en moet worden verstrekt op een wijze die nuttig en begrijpelijk is, zodat individuen hun rechten effectief kunnen beschermen. - Het recht van een gebruiker van een AI-systeem om door een mens te worden bijgestaan wanneer een AI-systeem gebruikt wordt voor interactie met personen, in het bijzonder in de context van overheidsdiensten.	- Individuen moeten duidelijk worden geïnformeerd over hun recht om door een mens te worden bijgestaan wanneer een AI-systeem wordt gebruikt dat hun rechten kan aantasten of hen op vergelijkbare wijze in aanzienlijke mate kan beïnvloeden, in het bijzonder in de context van overheidsdiensten, en over de wijze waarop zij om dergelijke bijstand kunnen verzoeken. Lidstaten moeten de ontwikkelaars en gebruikers van AI-systemen ertoe verplichten adequate communicatie te verstrekken. - Wanneer het gebruik van AI-systemen negatieve gevolgen dreigt te hebben voor mensenrechten, democratie of de rechtsstaat, moeten de lidstaten aan de ontwikkelaars en gebruikers van AI vereisten stellen inzake traceerbaarheid en informatieverstrekking. - De lidstaten moeten alle relevante informatie over AI-systemen (met inbegrip van hun werking, optimaliseringsfunctie, onderliggende logica, type gebruikte data) die worden gebruikt bij het verlenen van overheidsdiensten, openbaar en toegankelijk maken, waarbij legitieme belangen zoals openbare veiligheid of intellectuele eigendomsrechten worden beschermd, terwijl de mensenrechten evenwel volledig worden geëerbiedigd.
GEGEVENSBESCHERMING EN RECHT OP PRIVACY	- Het recht op eerbiediging van het privé-leven en van het familie- en gezinsleven en op bescherming van persoonsgegevens **(art. 8 EVRM).** - Het recht op fysieke, psychologische en morele integriteit in het licht van op AI gebaseerde profilering en emotie-/persoonlijkheidsherkenning. - Alle rechten die zijn neergelegd in Verdrag 108+ en in de gemoderniseerde versie daarvan, in het bijzonder met betrekking tot AI-profilering en het traceren van locaties.	- Lidstaten moeten ervoor zorgen dat het recht op privacy en gegevensbescherming wordt gewaarborgd gedurende de gehele levenscyclus van AI-systemen die zij inzetten, of die door private actoren worden ingezet. - Lidstaten moeten maatregelen nemen om individuen doeltreffend te beschermen tegen door AI gedreven massasurveillance, zoals door middel van biometrische herkenningstechnologie op afstand of andere door AI ondersteunde tracking-technologie. - Bij de aanschaf of implementatie van AI-systemen moeten Lidstaten de negatieve gevolgen voor het recht op privacy en gegevensbescherming en voor het bredere recht op eerbiediging van het privé- en gezinsleven beoordelen en beperken. Van bijzonder belang is de proportionaliteit van het invasieve karakter van het systeem in het licht van het legitieme doel dat het moet dienen, alsmede de noodzaak om dat doel te bereiken. - Lidstaten moeten passende waarborgen voor grensoverschrijdende datastromen invoeren om ervoor te zorgen dat de regels inzake gegevensbescherming niet worden omzeild.

	MATERIËLE RECHTEN	BELANGRIJKSTE VERPLICHTINGEN
VERANTWOORDING EN VERANTWOORDELIJKHEID	- Het recht op een daadwerkelijk rechtsmiddel in geval van schending van rechten en vrijheden **(art. 13 EVRM)**. - Dit moet ook het recht op effectieve en toegankelijke rechtsmiddelen omvatten wanneer de ontwikkeling of het gebruik van AI-systemen door private of publieke entiteiten ongerechtvaardigde schade veroorzaakt of inbreuk maakt op wettelijk beschermde rechten van een individu.	- De lidstaten moeten ervoor zorgen dat in het kader van hun respectieve nationale rechtsstelsels doeltreffende rechtsmiddelen beschikbaar zijn, onder meer voor burgerrechtelijke en strafrechtelijke aansprakelijkheid, en dat toegankelijke beroepsmogelijkheden worden ingesteld voor personen wier rechten negatief worden beïnvloed door de ontwikkeling of het gebruik van AI-toepassingen. - Lidstaten moeten mechanismen voor publiek toezicht instellen voor AI-systemen die mogelijk inbreuk maken op mensenrechten, democratie of de rechtsstaat. - Lidstaten moeten ervoor zorgen dat de ontwikkelaars en gebruikers van AI-systemen (1) de mogelijke negatieve gevolgen van AI-systemen voor mensenrechten, democratie en de rechtsstaat identificeren, documenteren en rapporteren; en (2) adequate risicobeperkende maatregelen nemen om verantwoordelijkheid en aansprakelijkheid voor eventuele schade te garanderen. - Lidstaten moeten maatregelen nemen om ervoor te zorgen dat overheden te allen tijde de door private actoren gebruikte AI-systemen kunnen controleren, om na te gaan of zij in overeenstemming zijn met de bestaande wetgeving en om private actoren verantwoordelijk te kunnen houden.
DEMOCRATIE	- Het recht op vrijheid van meningsuiting, van vergadering en van vereniging **(art. 10 en 11 EVRM).** - Het recht te stemmen en verkozen te worden, het recht op vrije en eerlijke verkiezingen, en in het bijzonder het universeel, gelijkwaardig en vrij stemrecht, met inbegrip van gelijke kansen en de vrijheid van de kiezers een mening te vormen. In dit verband mogen mensen niet worden onderworpen aan bedrog of manipulatie. - Het recht op (diverse) informatie, vrij discours en toegang tot pluraliteit van ideeën en zienswijzen. - Het recht op behoorlijk bestuur.	- Lidstaten moeten passende maatregelen nemen om het gebruik of misbruik van AI-systemen voor onrechtmatige inmenging in verkiezingsprocessen, gepersonaliseerde politieke targeting zonder passende transparantie-, verantwoordelijkheids- en verantwoordingsmechanismen, of meer in het algemeen voor het beïnvloeden van stemgedrag van kiezers of voor het manipuleren van de publieke opinie, tegen te gaan. - Lidstaten moeten strategieën vaststellen en maatregelen nemen ter bestrijding van desinformatie en online haatzaaien om eerlijke en diverse informatie te waarborgen. - Lidstaten moeten hun procedures voor overheidsopdrachten onderwerpen aan juridisch bindende voorwaarden die een verantwoord gebruik van AI in de overheidssector garanderen door waarborging van de naleving van de bovengenoemde beginselen, waaronder transparantie, billijkheid, verantwoordelijkheid en toerekenbaarheid. - Lidstaten moeten maatregelen nemen om de digitale geletterdheid en digitale vaardigheden in alle groepen van de bevolking te vergroten. Hun onderwijscurricula moeten worden aangepast om een cultuur van verantwoorde innovatie te promoten waarin mensenrechten, de democratie en de rechtsstaat worden geëerbiedigd.

RECHTSSTAAT

MATERIËLE RECHTEN

- Het recht op een eerlijk proces en een eerlijke rechtsgang **(art. 6 EVRM)**. Dit moet ook de mogelijkheid omvatten om inzicht te krijgen in en bezwaar te maken tegen AI-gedreven beslissingen in het kader van rechtshandhaving of justitie, met inbegrip van het recht op toetsing van een dergelijke beslissing door een mens.

- Het recht op rechterlijke onafhankelijkheid en onpartijdigheid, en het recht op rechtsbijstand.

- Het recht op een doeltreffende voorziening in rechte **(art. 13 EVRM)**, ook in geval van schade of schending van de mensenrechten in de context van AI-systemen.

BELANGRIJKSTE VERPLICHTINGEN

- Lidstaten moeten ervoor zorgen dat de AI-systemen die in het kader van justitie en rechtshandhaving worden gebruikt, in overeenstemming zijn met de essentiële eisen van het recht op een eerlijk proces. Daartoe moeten zij de kwaliteit en de veiligheid van rechterlijke beslissingen en data, alsmede de transparantie, onpartijdigheid en billijkheid van de dataverwerkingsmethoden waarborgen. Met het oog hierop moeten waarborgen worden ingevoerd voor de toegankelijkheid en de uitlegbaarheid van de dataverwerkingsmethoden, met inbegrip van de mogelijkheid van externe audits.
- Lidstaten moeten ervoor zorgen dat doeltreffende rechtsmiddelen beschikbaar zijn en dat toegankelijke verhaalsmechanismen worden ingesteld voor personen wier rechten worden geschonden door de ontwikkeling of het gebruik van AI-systemen in contexten die voor de rechtsstaat van belang zijn.
- Lidstaten moeten individuen zinvolle informatie verstrekken over het gebruik van AI-systemen in de overheidssector telkens wanneer dit het leven van individuen aanmerkelijk kan beïnvloeden. Dergelijke informatie dient in het bijzonder te worden verstrekt wanneer AI-systemen worden gebruikt op het gebied van justitie en rechtshandhaving, zowel wat betreft de rol van AI-systemen binnen het proces, als wat betreft het recht om de besluiten die door AI-systemen geïnformeerd of genomen, aan te vechten.
- Lidstaten moeten ervoor zorgen dat het gebruik van AI-systemen geen afbreuk doet aan de beslissingsbevoegdheid van rechters of de onafhankelijkheid van de rechterlijke macht, en dat alle rechterlijke beslissingen onderworpen zijn aan adequaat menselijk toezicht.

AANVULLENDE OVERWEGINGEN INZAKE BEGINSELEN, RECHTEN EN VERPLICHTINGEN

Er zijn enkele bijkomende factoren die moeten worden meegewogen bij de mogelijke invoering van nieuwe rechten en verplichtingen in een toekomstig juridisch kader inzake AI-systemen. Ten eerste moeten deze rechten en plichten noodzakelijk, nuttig en proportioneel zijn ten aanzien van het doel om burgers te beschermen tegen de negatieve gevolgen van AI-systemen voor mensenrechten, de democratie en de rechtsstaat. Tegelijkertijd moet gezorgd worden voor een rechtvaardige en billijke verdeling van de voordelen van AI. Deze afweging van risico's en voordelen moet alomvattend zijn en moet een bewustzijn van het evenwicht tussen de legitieme belangen die op het spel staan, inhouden. Een op risico's en voordelen gebaseerde aanpak moet ook onderscheid maken tussen verschillende risiconiveaus en moet daarmee rekening houden wanneer wetgevende maatregelen geformuleerd en overeengekomen worden.

Belangrijkste elementen van een risicogebaseerde en van voordelen bewuste aanpak:

- Houd rekening met de **gebruikscontext** en de **potentiële impact** van de AI-technologie
- Houd rekening met het **toepassingsgebied** en **de betrokken belanghebbenden**
- **Beoordeel de risico's regelmatig** en **systematisch**, en **stem eventuele risicobeperkende maatregelen af** op deze risico's
- **Optimaliseer de maatschappelijke voordelen** van AI-innovatie **door gerichte risico-gebaseerde wetgevende maatregelen**

Nationale autoriteiten moeten een centrale rol spelen bij de systematische beoordeling of nationale wetgeving in overeenstemming is met het streefdoel de ontwikkeling en het gebruik van AI in lijn te brengen met mensenrechten, democratie en de rechtsstaat, en om eventuele lacunes in de wetgeving op te sporen. Tevens moeten nationale mechanismen voor de controle van en het toezicht op AI-systemen bescherming bieden tegen schadelijke gevallen van niet-naleving. Ten slotte, aangezien private actoren steeds vaker noodzakelijke digitale infrastructuur voor de publieke sector leveren die van invloed is op het algemeen belang, hebben zij de verantwoordelijkheid om het ontwerp, de ontwikkeling en de inzet van hun technologieën af te stemmen op deze beginselen en prioriteiten.

06 LANDSCHAP VAN JURIDISCHE INSTRUMENTEN

INTERNATIONALE WETGEVENDE KADERS

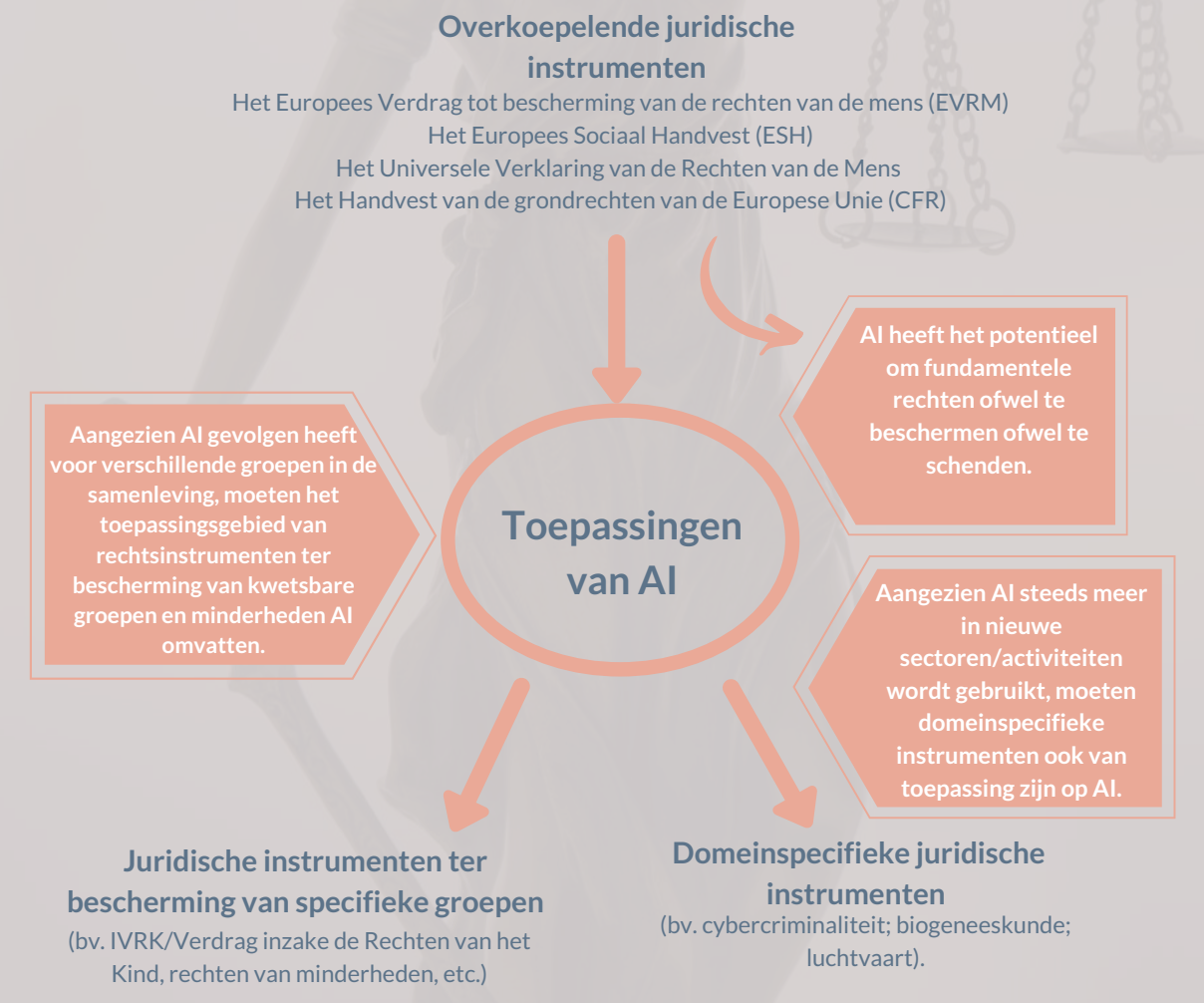
Momenteel zijn er geen internationale wetten die specifiek betrekking hebben op AI - of op geautomatiseerde besluitvorming - maar een aantal bestaande wetgevende kaders zijn relevant. In het bijzonder (zoals hierboven samengevat):

- Het Europees Verdrag voor de Rechten van de Mens (EVRM)
- Het Europees Sociaal Handvest (ESH)
- Het Universele Verklaring van de Rechten van de Mens
- Het Handvest van de grondrechten van de Europese Unie (CFR)

In deze juridische instrumenten zijn de fundamentele rechten van de mens vastgelegd, waarvan er vele relevant zijn voor toepassingen van AI, zoals bijvoorbeeld: het recht op non-discriminatie en het recht op privacy.

Ook bestaan er juridische instrumenten die de rechten van mensen met betrekking tot bepaalde sectoren en/of activiteiten vastleggen, waaronder cybercriminaliteit, biogeneeskunde en luchtvaart. Aangezien AI in toenemende mate gebruikt wordt in diverse sectoren en op manieren die steeds meer delen van ons leven beïnvloeden, wordt het ook relevanter voor elk van deze rechtsgebieden.

AI is ook relevant voor juridische instrumenten die dienen om kwetsbare groepen of minderheden te beschermen. Hoewel er dus geen specifiek juridisch mechanisme voor AI bestaat, zijn veel van de bestaande juridische mechanismen relevant voor de wijze waarop AI wordt ontwikkeld en ingezet.



HUIDIGE SOFT LAW BENADERINGEN

Momenteel zijn de voornaamste beleids- en reguleringsactiviteiten voor AI gebaseerd op "soft law"-benaderingen. Het verschil tussen harde wetgeving en “soft law” wordt hieronder toegelicht.

De laatste jaren zijn er steeds meer richtsnoeren en beginselen gekomen voor ethisch handelen met betrekking tot AI. Deze zijn meestal bedoeld om de betrouwbaarheid aan te tonen van de manier waarop AI wordt ontwikkeld en ingezet. Dergelijke richtsnoeren of beginselen zijn ontwikkeld door organisaties uit de private sector, de academische wereld en de publieke sector. In veel gevallen heeft de ontwikkeling van interne richtsnoeren en goede praktijken gediend als een argument tegen de noodzaak van harde wetgeving met betrekking tot AI of meer gecentraliseerde regulering van AI. Veel organisaties die beginselen of richtsnoeren voor ethische AI hebben voorgesteld, hebben sterk gepleit voor zelfregulering.

Vrijwillige gedragscodes binnen organisaties die gebruik maken van AI kunnen een belangrijke rol spelen bij het vormgeven van de organisatiecultuur en leiden tot nuttige effecten in de praktijk. Bovendien hebben zij als voordeel dat zij flexibel en aanpasbaar zijn, onmiddellijk kunnen worden geïmplementeerd, een breder publiek aanspreken en snel kunnen worden herzien en gewijzigd. Zij worden echter ook bekritiseerd omdat zij symbolisch en grotendeels retorisch zouden zijn.

Er is een zekere mate van consistentie in de beginselen die in de bestaande richtsnoeren naar voren worden gebracht. Transparantie wordt bijvoorbeeld stelselmatig benadrukt. Er is daarentegen een gebrek aan consistentie omtrent de praktische begeleiding. Dit leidt tot zeer uiteenlopende benaderingen en verschillende opvattingen over wat ethisch vereist is of hoe AI moet worden gereguleerd. Bovendien is er, hoewel er geen tekort is aan gedragscodes of richtsnoeren voor ethische AI, over het algemeen een gebrek aan verantwoording en transparantie met betrekking tot de handhaving van deze codes. Handhaving door middel van interne comités of beoordelingspanels wordt bekritiseerd omdat deze niet transparant of doeltreffend genoeg zou zijn.

Er valt dan ook veel te zeggen voor een combinatie van vrijwillige “soft law” initiatieven in combinatie met bindende regelgeving.

Harde wetgeving

- Juridisch bindende instrumenten
- Wettelijke sancties
- Afdwingbaar in rechte

Soft law

- Niet-bindende aanbevelingen, richtsnoeren, certificeringen of verklaringen die gemeenschappelijke beginselen en ‘best practices’ vastleggen
- Vaak multi-interpretabel
- Geen wettelijke sancties

NATIONALE WETGEVENDE INSTRUMENTEN

Internationaal is er groeiende belangstelling voor het ontwikkelen van benaderingen om AI te ‘managen’ of te reguleren. Soft law-benaderingen overheersen. Uit een raadpleging van CAHAI-leden bleek dat:

- 30 lidstaten en 4 waarnemende landen strategieën en beleid met betrekking tot AI-systemen hebben;
- 1 lidstaat een vrijwillig AI-certificeringsprogramma gelanceerd heeft;
- 2 lidstaten formeel hun goedkeuring gehecht hebben aan internationale of Europese niet-bindende ethische kaders voor AI;
- 12 lidstaten en 4 waarnemende landen een of meer wetgevende of ethische instrumenten hebben aangenomen.

Deze initiatieven werden geleid door verschillende instellingen, waaronder nationale raden, comités, gespecialiseerde AI-overheidsinstellingen en overheidsinstanties.

Wat de ontwikkeling van harde wetgeving betreft, bleek uit de raadpleging van CAHAI-leden dat:

- 4 lidstaten specifieke wettelijke kaders aangenomen hebben voor AI bij het testen en gebruiken van autonome voertuigen (zelfrijdende auto's);
- 2 lidstaten bezig zijn met de ontwikkeling van wetgevende kaders met betrekking tot het gebruik van AI bij werving en selectie en geautomatiseerde besluitvorming door overheidsinstanties.

DE ROL VAN PRIVATE ACTOREN

Private actoren (bv. bedrijven) hebben het domein van 'AI-ethiek' in belangrijke mate gevormd, onder meer via vrijwillige gedragscodes. In sommige gevallen hebben private actoren ook gepleit voor een regelgevend kader om de rechtszekerheid rond AI te vergroten.

Het is duidelijk dat de private actoren een belangrijke rol spelen. De verantwoordelijkheid van private actoren om bij hun activiteiten, producten en diensten mensenrechten te respecteren, is vastgelegd in de VN-richtsnoeren voor bedrijven en mensenrechten.

Indien een nieuwe regelgevende aanpak wordt toegepast, zullen de betrokkenheid en de medewerking van private actoren van cruciaal belang zijn voor de ontwikkeling van sectorale soft law. Dit zal belangrijk zijn om de tenuitvoerlegging van de wetgeving in de specifieke context te ondersteunen en aan te vullen (bijvoorbeeld via sectorspecifieke richtsnoeren of certificeringsregelingen).

Een doeltreffend regelgevingskader voor AI zal nauwe samenwerking vereisen tussen alle belanghebbenden, met inbegrip van staten, overheidsinstanties, de het maatschappelijk middenveld en bedrijven, zodat rekening wordt gehouden met uiteenlopende belangen en perspectieven.

HUIDIGE BEPERKINGEN

Veel van de juridische instrumenten die momenteel worden gebruikt om aspecten van AI te reguleren, zijn ontwikkeld voordat AI-systemen gangbaar werden. Als zodanig kunnen zij ontoereikend zijn om de diverse effecten en risico's van AI aan te pakken.

Soft law benaderingen zijn niet-bindend en berusten op vrijwillige naleving, wat kan leiden tot uiteenlopende praktijken en resultaten. Bovendien kunnen de uiteenlopende benaderingen van organisaties op het gebied van soft law leiden tot symbolische of cosmetische verplichtingen inzake ethische AI. Niettemin kan veel werk dat nu wordt verricht op het gebied van normen en certificering, toekomstige wettelijke mechanismen ondersteunen.

Daarnaast zijn er belangrijke beginselen die momenteel niet wettelijk zijn gewaarborgd bij het gebruik van AI. Bijvoorbeeld de noodzaak om te zorgen voor voldoende menselijke controle en toezicht, en de doeltreffende transparantie over en uitlegbaarheid van AI-systemen. Er is een gebrek aan juridische instrumenten om deze belangrijke technische factoren van AI aan te pakken.

Hoewel de huidige juridische mechanismen tot op zekere hoogte individuele rechten beschermen, worden de maatschappelijke dimensies van de risico's van AI nog niet voldoende geadresseerd (bv. risico's voor verkiezingen of democratische instellingen). De bescherming van de democratie en de rechtsstaat vereist betrokkenheid van, en toezicht door, de overheid bij een verantwoorde ontwikkeling en gebruik van AI-systemen.

Ten slotte leiden de huidige lacunes in de wetgeving tot onzekerheid en ambiguïteit rondom AI. Dit is van belang voor de ontwikkelaars en gebruikers van AI en voor de samenleving in het algemeen. Onzekerheid op dit gebied kan een belemmering vormen voor de voordelen van AI-innovatie en kan in de weg staan aan belangrijke innovatie die ten goede zou kunnen komen van burgers en de gemeenschappen waarin zij leven.

TOEKOMSTIGE BEHOEFTEN EN MOGELIJKHEDEN

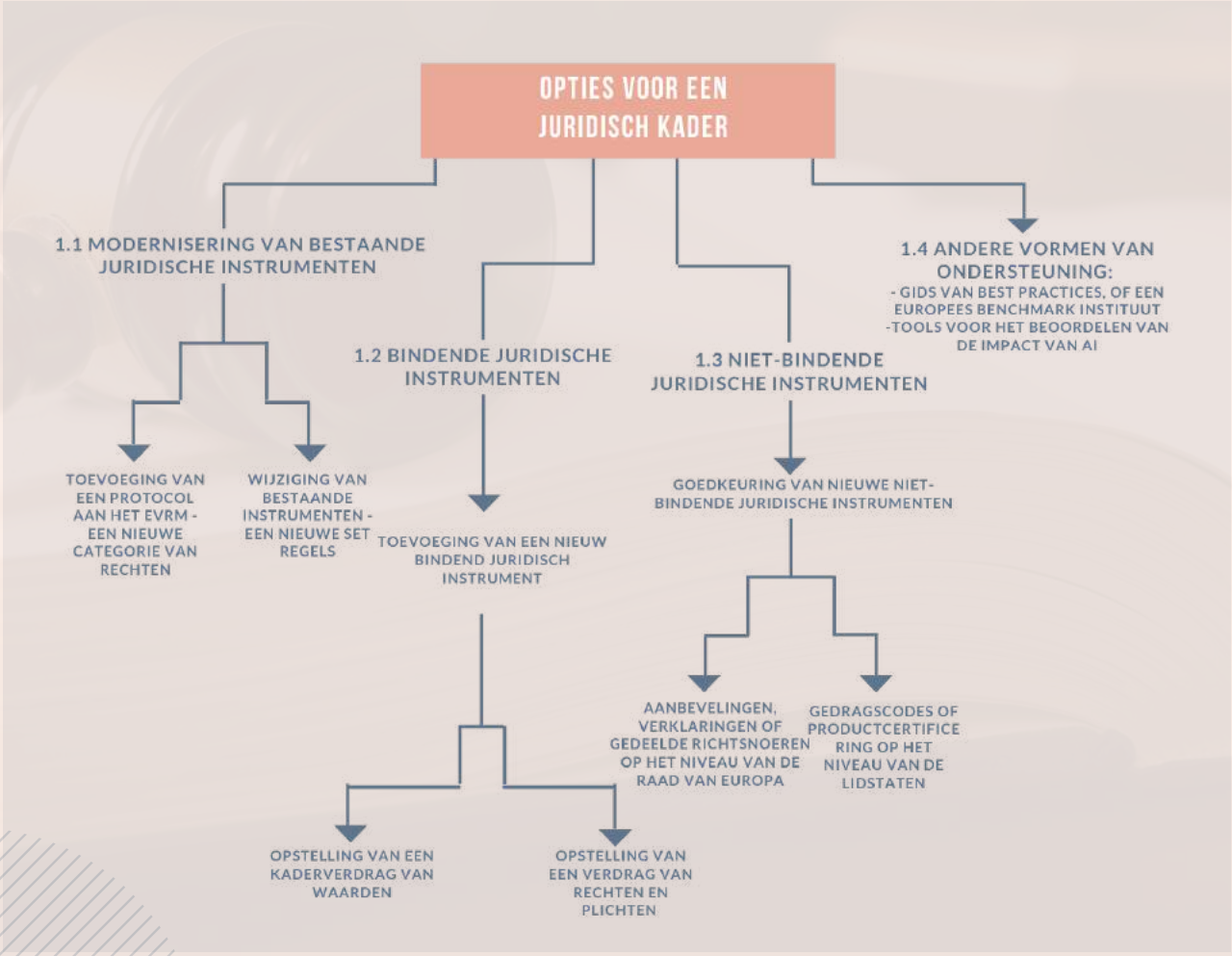
Toekomstige strategieën voor regelgeving moeten de hierboven uiteengezette beperkingen aanpakken. Zij moeten sectoroverschrijdend zijn en bindende bepalingen bevatten om te zorgen voor bescherming van mensenrechten, democratie en de rechtsstaat, en een meer alomvattende bescherming te garanderen. Dit zou een aanvulling kunnen vormen op bestaande sectorspecifieke voorschriften.

De ontwikkeling van een juridisch bindend instrument op basis van de normen van de Raad van Europa - mocht het Comité van Ministers deze optie ondersteunen - zou het initiatief van de Raad van Europa een unieke positie geven ten opzichte van andere internationale initiatieven, die ofwel gericht zijn op het uitwerken van een ander soort instrument, ofwel een ander toepassingsgebied of een andere achtergrond hebben.

OPTIES VOOR EEN JURIDISCH KADER

Er zijn verschillende manieren waarop de Raad van Europa zou kunnen besluiten regels op te stellen voor AI ter bescherming van mensenrechten, democratie en de rechtsstaat. Elke benadering heeft voor- en nadelen in termen van te verwachten resultaten.

Er zijn twee belangrijke verschillen om in overweging te nemen. Het eerste is het verschil tussen bindende en niet-bindende juridische instrumenten, waarbij Staten al dan niet gebonden zijn aan de regels waartoe de Raad besluit. Het tweede is de vraag in welke mate bestaande instrumenten geconsolideerd en gemoderniseerd moeten worden en in hoeverre volledige nieuwe instrumenten moeten worden gecreëerd. Zie de onderstaande grafiek voor een overzicht van deze benaderingen en waar in dit deel meer informatie te vinden is.



1.1: Modernisering van bestaande bindende juridische instrumenten

Een van de opties die worden overwogen is de wijziging van bestaande regels in de context van AI. Bijvoorbeeld door het toevoegen van een protocol (een reeks rechten) aan het bestaande Europees Verdrag tot bescherming van de Rechten van de Mens. Een aanvullend protocol zou een krachtige uiting van ondersteuning van de bescherming van de mensenrechten, de democratie en de rechtsstaat in de context van AI zijn door de lidstaten, maar zou op zichzelf niet de mogelijkheid bieden om meer specifieke vereisten of normen vast te stellen. Aanvullende protocollen zijn slechts bindend voor staten die ze ratificeren, waardoor het overzicht gefragmenteerder kan worden. Het Europees Hof voor de Rechten van de Mens is bovendien reeds overbelast met zaken.

Een andere mogelijkheid is dat de Raad besluit bestaande instrumenten (die bestaan uit een set van regels) te wijzigen om er de door AI naar voren gebrachte overwegingen in op te nemen. Twee bestaande instrumenten die in die zin gewijzigd zouden kunnen worden, zijn het Verdrag van Boedapest inzake de bestrijding van strafbare feiten verbonden met elektronische netwerken, en Verdrag 108+, dat de verwerking van persoonsgegevens over natuurlijke personen waarborgt. Een voordeel van deze aanpak is dat er bestaande capaciteit is voor het toezicht op en de handhaving van de reeds bestaande regels. Een nadeel van deze aanpak is echter dat het lastig is de bestaande instrumenten voldoende aan te passen. De uitdagingen op het gebied van cybercriminaliteit en gegevensbescherming zijn verwant, maar niet identiek, aan de uitdagingen die AI oproept, zoals de verantwoordelijkheid voor en de uitlegbaarheid van geautomatiseerde systemen.

Een laatste mogelijkheid is om deze twee opties te combineren om zo de nadelen van elk van beide te vermijden. Door een protocol toe te voegen kunnen algemene beginselen en waarden worden vastgesteld, en door bestaande instrumenten te wijzigen kunnen de verplichtingen van staten om deze beginselen in de praktijk te beschermen, nader worden uitgewerkt, terwijl tegelijkertijd voldoende capaciteit gegarandeerd wordt om hierop toezicht te houden. De vraag is of een gecombineerde aanpak niet te traag en omslachtig zou zijn, gezien het hoge tempo waarin AI wordt ontwikkeld en ingezet.

1.2: Goedkeuring van een nieuw bindend juridisch instrument

Een alternatieve aanpak zou kunnen zijn om een geheel nieuwe reeks bindende regels specifiek voor AI te ontwikkelen en aan te nemen. Dit kan op twee manieren: een *verdrag* of een *kaderverdrag*. Net als bij het onderscheid tussen *protocollen* en *wijziging van bestaande instrumenten met nieuwe regels* hierboven, worden in een *kaderverdrag* algemene beginselen en actieterreinen vastgesteld, terwijl in een *verdrag* een specifieke materie concreet wordt geregeld door het creëren van rechten en verplichtingen. Als verdragen hebben zij echter dezelfde internationaalrechtelijke status. Laten we ze een voor een bekijken.

Een kaderverdrag zou kunnen voorzien in algemene beginselen en kernwaarden die gerespecteerd moeten worden bij het ontwerp en gebruik van AI-systemen, maar het zou de lidstaten een aanzienlijke discretionaire vrijheid laten om te bepalen hoe deze beginselen en waarden in de praktijk moeten worden toegepast. Nadat het Kaderverdrag tot stand is gekomen, zouden de ondertekenaars van het verdrag kunnen besluiten meer gedetailleerde protocollen en specifieke bepalingen te creëren. Deze aanpak zou zeer geschikt kunnen zijn voor de snelle ontwikkeling van AI en de nieuwe ethische vraagstukken die het oproept. Een *kaderverdrag* zou kunnen voorzien in beginselen en regels voor de ontwikkeling van AI, alsook in specifieke richtsnoeren over hoe toezicht en samenwerking tussen landen kan worden gegarandeerd. Tussen leden van de Raad van Europa bestaan dergelijke overeenkomsten al voor de bescherming van nationale minderheden, en de bescherming van mensen in de context van medische behandeling en onderzoek - wat opmerkelijk is omdat beide kwesties enige overlap vertonen met de potentiële schade van AI-systemen. Kenmerkend voor kaderverdragen is echter dat zij alleen algemene verplichtingen voor staten vastleggen en geen concrete rechten voor mensen, zodat staten speelruimte hebben bij de wijze waarop zij de beginselen implementeren.

Verdragen kunnen een meer alomvattende regeling mogelijk maken. In het geval van AI zou een verdrag de rechten en plichten kunnen vaststellen die mensenrechten, de democratie en de rechtsstaat waarborgen, en als gevolg daarvan een betere rechtsbescherming bieden aan mensen. Een verdrag zou staten aanmoedigen om snel te handelen en relevante nationale wetten in te voeren, en het zou een gelijk speelveld creëren voor verantwoorde, betrouwbare AI-producten, zelfs over nationale grenzen heen.

De weg van het verdrag houdt echter het risico in dat het al te rigide kan zijn en nieuwe toepassingen van AI die de samenleving ten goede kunnen komen zou kunnen belemmeren. Niettemin zou een concrete reeks internationaal bindende regels alle betrokkenen rechtszekerheid bieden, degelijke bescherming bieden aan wie negatieve gevolgen van AI ondervindt, en de basis leggen voor daadwerkelijk verantwoorde ontwikkeling van AI.

Ongeacht de keuze voor een *kaderverdrag* of een *verdrag*, zijn de geadresseerden van dit instrument (d.w.z. degenen voor wie de regels in de eerste plaats bedoeld zijn) staten, die, door het verdrag formeel aan te nemen, ermee instemmen om volgens het internationaal recht aan de voorwaarden ervan gebonden te zijn. Het tijdschema voor de goedkeuring van een verdrag is echter onduidelijk, en zelfs staten die in de Raad van Europa voor het verdrag hebben gestemd, zijn niet verplicht het formeel aan te nemen. Bovendien is het van belang ervoor te zorgen dat andere actoren, zoals naties buiten Europa, gelijkwaardige regels aannemen, omdat anders de internationale regels en normen voor AI gefragmenteerd zouden kunnen raken.

1.3 Niet-bindende juridische instrumenten

Niet-bindende instrumenten of "soft law" instrumenten hebben niet dezelfde internationale rechtskracht, maar kunnen niettemin een richtinggevende rol spelen voor Staten en andere actoren. Hoewel soft law op zichzelf niet kan garanderen dat AI gericht is op mensenrechten, democratie en de rechtsstaat, kan het daartoe bijdragen en heeft het als voordeel dat het flexibel, aanpasbaar en snel toepasbaar is. Niet-bindende juridische instrumenten kunnen worden vastgesteld op het niveau van de Raad van Europa worden vastgesteld of door de lidstaten worden goedgekeurd. Deze instrumenten sluiten elkaar niet uit, maar laten we ze stuk voor stuk bekijken.

Een ruim soft law instrument op het niveau van de Raad van Europa zou een aanbeveling of een verklaring kunnen zijn, hetzij als een op zichzelf staand document, hetzij als aanvulling op een van de hierboven besproken bindende instrumenten. Een andere optie is het creëren van begeleidende documenten of richtsnoeren die licht werpen op de implicaties van AI voor mensenrechten, democratie en de rechtsstaat. Deze documenten zouden ontwikkeld kunnen worden in samenwerking met alle betrokken partijen, met inbegrip van vertegenwoordigers van de overheid, de private sector, het maatschappelijk middenvelden de academische wereld, en zouden "evolutief" zijn, en over de tijd geüpdatet kunnen worden om rekening te houden met nieuwe ontwikkelingen.

Op het niveau van de lidstaten zouden "soft law" instrumenten de vorm kunnen aannemen van richtsnoeren, gedragscodes, labels, merktekens of certificieringszegels voor AI-producten. Deze voorbeelden van soft law zouden kunnen worden opgenomen in de bestuurs-, aanbestedings- en auditpraktijken van organisaties zoals private ondernemingen. Hoewel deze vorm van "zelfregulering" een aanvulling kan vormen op andere beginselen en regels, mag zij niet in de plaats komen van de verplichtingen van de lidstaten om mensenrechten, democratie en de rechtsstaat actief te beschermen.

1.4 Andere vormen van ondersteuning

Naast bindende en niet-bindende juridische instrumenten kunnen aan de lidstaten en andere actoren ook andere vormen van steun worden verleend. Dit omvat de mogelijkheid van best practices om positieve actie te helpen sturen. De oprichting van een "Europees benchmark-instituut" zou een doeltreffende manier kunnen zijn om dergelijke best practices vast te stellen en te bepalen hoe zij moeten worden ondersteund, en daarover een consensus te bereiken. Bovendien zou het creëren van een model of instrument waarmee de impact van AI op het niveau van de Raad van Europa kan worden beoordeeld, kunnen helpen om de toepassing van normen en waarden inzake AI in het hele continent op hetzelfde niveau te brengen.

Samenvattend kunnen we stellen dat elke aanpak om er daadwerkelijk voor te zorgen dat AI de mensenrechten, de democratie en de rechtsstaat waarborgt, waarschijnlijk een combinatie zal vergen van de hier geschetste horizontale (bindende en niet-bindende) benaderingen en meer sectorspecifieke beginselen, normen en vereisten.

07 PRAKTISCHE MECHANISMEN TER ONDERSTEUNING VAN HET JURIDISCHE KADER

Welke praktische mechanismen zijn beschikbaar om de doeltreffendheid van het juridisch kader te helpen ondersteunen, naleving te garanderen en best practices te bevorderen? We zullen een aantal antwoorden op deze vragen onderzoeken door te kijken naar de rol van de mechanismen en de relevante actoren, en enkele voorbeelden geven van mechanismen om a) naleving te ondersteunen en b) follow-up activiteiten te ondersteunen.

DE ROL VAN HANDHAVINGSMECHANISMEN

Er bestaan diverse praktische mechanismen om de handhaving van regels te ondersteunen en te garanderen, waaronder "due diligence" inzake mensenrechten, effectbeoordelingen, certificering en normen, auditing en monitoring, en zelfs "regulatory sandboxes". Deze mechanismen ondersteunen de naleving van het juridisch kader, maar bieden ook extra voordelen, zoals meer transparantie en vertrouwen. Zij bevorderen ook best practices binnen en tussen sectoren, zoals een reflectieve en anticiperende beoordeling van AI-systemen, vanaf de vroegste fases van het projectontwerp tot het gebruik en de monitoring ervan na de invoering ervan.

Het juridisch kader stelt overkoepelende vereisten vast voor de wijze waarop deze mechanismen moeten worden ontwikkeld. Het kan bijvoorbeeld bepalen dat handhavingsmechanismen moeten evolueren, parallel met de ontwikkeling en het gebruik van een systeem, om rekening te houden met eventuele wijzigingen in de werking ervan.

Hoewel het juridisch kader op beginselen gebaseerde vereisten moet vaststellen voor de wijze waarop handhavingsmechanismen moeten worden ontwikkeld, moet het de verantwoordelijkheid van de lidstaten blijven om deze mechanismen te implementeren op basis van de bestaande rol van de lokale instellingen en de wetgevingscultuur.

Van Naleving naar Zekerheid

Praktische mechanismen kunnen ook worden gebruikt om zekerheid te verschaffen aan relevante exploitanten of gebruikers, alsook om best practices te bevorderen. Dit kader breidt de rol van praktische mechanismen uit tot buiten het perspectief van louter naleving en helpt bij het bevorderen van een ecosysteem van zekerheid dat talloze voordelen biedt, waaronder:

- ondersteuning van interne **reflectie** en **overleg** door het aanreiken van praktische tools voor de evaluatie van het ontwerp, de ontwikkeling en het gebruik van AI-gedreven systemen of producten, waarbij een dynamische aanpak wordt gehanteerd die samen met het systeem evolueert (bv. monitoring van veranderingen in het gedrag van het systeem na de invoering)
- facilitering van **transparante communicatie** tussen ontwikkelaars, toezichthouders, exploitanten en gebruikers, en overige belanghebbenden
- ondersteuning van **documentatie-**processen (of verslaglegging) om de verantwoording te waarborgen (bv. audits)
- het opbouwen van **betrouwbaarheid en vertrouwen** door best practices te bevorderen en toe te passen (bv. normen of certificeringsregelingen).

DE ROL VAN VERSCHILLENDE ACTOREN

In grote lijnen kunnen aan de hand van de volgende drie categorieën actoren worden geïdentificeerd die elk op complementaire wijze bij kunnen dragen tot het garanderen van de naleving van de nationale regelgeving.

ACTOR	ROL VAN DE ACTOR
Toezichthouders op systemen	Onafhankelijke toezichtsorganen, zoals deskundigencommissies, sectorale toezichthouders of auditors uit de private sector, moeten duidelijk geïdentificeerde belanghebbenden die de gevolgen ondervinden van de praktische toepassing van AI, vertegenwoordigen en aan hen verantwoording afleggen. Er mag evenwel niet worden verwacht dat hun takenpakket alle op AI-systemen omvat.
Ontwikkelaars van systemen	Ontwikkelaars in de private en publieke sector kunnen de naleving ondersteunen door beleidsmaatregelen te nemen die de zichtbaarheid vergroten van waar dergelijke technologieën worden ingezet (bv. door overheidscontracten te publiceren, of door openbare registers of meldingssystemen op te zetten). Gestandaardiseerde instrumenten voor interne audits en zelf-certificering hebben hun beperkingen, maar kunnen ook helpen.
Exploitanten en gebruikers van systemen	Goed geïnformeerde exploitanten en gebruikers van AI genereren vraag en kunnen deze inkoopmacht gebruiken om aanbieders en verkopers van AI-toepassingen ertoe aan te zetten het toekomstig wetgevend kader na te leven. Dit geldt in het bijzonder voor de overheidssector en zijn aanzienlijke inkoopmacht.

Tevens dient opgemerkt te worden dat veel AI-systemen, en de data waarop zij vertrouwen, in verschillende jurisdicties toegepast worden, waardoor het noodzakelijk is dat er adequate mechanismen voor informatie-uitwisseling en verslaglegging zijn ter ondersteuning van de taken van de betrokken actoren.

VOORBEELDEN VAN TYPES HANDHAVINGSMECHANISMEN

Er bestaat een grote verscheidenheid aan handhavingsmechanismen. Sommigen zullen het best werken in bepaalde contexten (bv. verschillende regelgevingsculturen) en afhankelijk zijn van de verschillende onderdelen van een AI-systeem die aan naleving onderworpen zijn (bv. kenmerken van de trainingsdata). Om te helpen bepalen welke mechanismen het best geschikt zijn voor welke context, moeten inclusieve en participatoire processen plaatsvinden met de relevante belanghebbenden.

Een aantal gemeenschappelijke kenmerken van doeltreffende praktische mechanismen zouden in een rechtskader kunnen worden gespecificeerd als in acht te nemen beginselen:

- **Dynamische (niet statische) beoordeling** aan het begin en gedurende de levenscyclus van een AI-project rekening houdend met de lopende besluitvorming
- Mechanismen moeten **technologisch aanpasbaar** zijn om toekomstbestendig te zijn
- De processen en outputs van de mechanismen moeten voor deskundigen en niet-deskundigen **toegankelijk en begrijpelijk** zijn om beroepsprocedures en verhaalsmogelijkheden te ondersteunen
- Er moet **onafhankelijk** toezicht zijn door de bevoegde instantie of partij (bv. auditor)
- **Bewezen** technische normen, certificeringen en praktijken moeten worden bevorderd en gebruikt

De volgende reeks mechanismen vormt een toolkit die aan veel van deze beginselen voldoet, maar ook ruimte biedt voor verfijning en innovatie op regelgevend gebied.

Due diligence mbt mensenrechten

Om ervoor te zorgen dat het ontwerp, de ontwikkeling en het gebruik van AI-systemen niet in strijd zijn met mensenrechten, is het van cruciaal belang dat organisaties de nodige zorgvuldigheid betrachten. Het gebruik van effectbeoordelingen is een praktische manier om nadelige gevolgen voor de mensenrechten die kunnen voortvloeien uit het gebruik van AI-systemen, te identificeren, voorkomen, beperken en verantwoorden. Effectief gebruik van effectbeoordelingen zal afhangen van de gebruikte sociaal-economische indicatoren en verzamelde data. Zo kan een effectbeoordeling bijvoorbeeld inzicht geven in welke impact een AI-systeem heeft op individueel welzijn, volksgezondheid, vrijheid, toegankelijkheid van informatie, sociaaleconomische ongelijkheid, milieuduurzaamheid, enzovoort.

Certificatie en kwaliteitslabels

Standaardenen certificeringsmechanismen worden op ruime schaal gebruikt als indicatoren voor veiligheid en kwaliteit en zouden kunnen worden uitgebreid tot AI-systemen (bv. certificering dat een bepaald systeem extensief geëvalueerd en getest is, op basis van industriële normen). Het toepassingsgebied van dergelijke regelingen kan betrekking hebben op de producten en systemen zelf of op de organisaties die verantwoordelijk zijn voor de ontwikkeling van de producten of systemen.

Auditing

Regelmatige audits door onafhankelijke, deskundige instanties die verantwoordelijk zijn voor het toezicht op een bepaalde sector (bv. gezondheidszorg) of een bepaald domein (bv. autonome voertuigen) kunnen de overgang naar een transparanter en verantwoordelijker gebruik van AI-systemen vergemakkelijken.

Regelluwe zones (regulatory sandboxes)

Het gebruik van regelluwe zones, beter gekend als regulatory sandboxes, biedt gemachtigde bedrijven de mogelijkheid om op AI-systemen, die niet door de huidige regelgeving worden beschermd, op een veilige en gecontroleerde manier (d.w.z. binnen een sandbox) te testen. Het gebruik van regulatory sandboxes kan de snelheid waarmee dergelijke producten of systemen op de markt komen, helpen verkorten en de kosten voor de organisatie verlagen, door innovatie op een gecontroleerde manier te ondersteunen.

Voortdurende, geautomatiseerde monitoring

Zodra AI-systemen zijn ingevoerd, moet de werking ervan voortdurend worden gecontroleerd om ervoor te zorgen dat de functionaliteit van het systeem blijft voldoen aan de verwachtingen. Het monitoringsproces kan worden geautomatiseerd om ervoor te zorgen dat afwijkingen in de functionaliteit van een op AI-systeem zo vroeg mogelijk worden opgespoord en aangepakt. Het gebruik van geautomatiseerde monitoring brengt echter ook risico's met zich zoals het mogelijk verlies van menselijk toezicht of de-skilling van professionele handhavingscontroleurs.

Het bevorderen en verplicht stellen van praktische mechanismen, zoals hierboven genoemd, dient plaats te vinden in samenhang met bredere, ondersteunende initiatieven, teneinde het potentieel ervan te maximaliseren. Zo zijn investeringen in digitale geletterdheid en de ontwikkeling van vaardigheden bij ontwikkelaars, beleidsmakers en wetgevers belangrijke voorwaarden voor de effectiviteit van een juridisch kader. Expertisecentra zijn goed gepositioneerd om deze bredere initiatieven te ondersteunen door een permanente dialoog, samenwerking en uitwisseling van best practices tussen actoren op nationaal en internationaal niveau te faciliteren.

OPVOLGMECHANISMEN

Naast bovengenoemde mechanismen zijn er diverse relevante opvolgmechanismen en -maatregelen. Een voorbeeld is het gebruik van onafhankelijke deskundigencommissies die belast kunnen worden met het toezicht op de tenuitvoerlegging en daadwerkelijk gebruik van juridische instrumenten (bv. een verdrag) of monitoring van de maatschappelijke gevolgen van AI-systemen. Zoals hierboven is opgemerkt, betekent het feit dat AI-systemen meerdere jurisdicties omvatten dat internationale samenwerking vereist zal zijn. Via netwerken tussen de verdragsluitende staten kan wederzijdse bijstand en samenwerking in straf- of burgerlijke zaken worden bevorderd.

08 CONCLUSIE

In deze brochure hebben wij getracht de belangrijkste elementen van de *Haalbaarheidsstudie* van CAHAI te introduceren, en hebben wij achtergrondinformatie verstrekt over de technische aspecten van AI en de verweven relatie tussen mensenrechten, democratie en de rechtsstaat. Wij hopen dat dit materiaal kan fungeren als een soort springplank voor een zinvolle reflectie op de vooruitzichten voor een juridisch kader voor AI, in overeenstemming met de bakens die de Raad van Europa al meer dan een halve eeuw uitzet op het vlak van mensenrechten, democratie en de rechtsstaat. Om deze transformerende en steeds krachtiger wordende technologie ten behoeve van zowel de burgers als de samenleving in brede zin op het juiste spoor te zetten, zijn goed geïnformeerde, visionaire beleidsvorming en zorgvuldige anticiperende reflectie vereist. De *Haalbaarheidsstudie* en deze ondersteunende introductie bieden de eerste stappen in die richting.

Nu het werk van CAHAI de fase van raadpleging van belanghebbenden en outreach ingaat, moet worden benadrukt dat de kwaliteit en het succes van deze belangrijke inspanning nu zullen afhangen van de wijsheid en de inzichten van een zo breed en inclusief mogelijke groep deelnemers. De democratische sturing van technologie en technologiebeleid vormt de kern van het mensgerichte en waardengedreven perspectief dat mensenrechten, democratie en de rechtsstaat in een centrale positie plaatst om de toekomst van AI beleid en digitale innovatie meer in het algemeen vorm te geven. Het is in feite alleen door middel van brede feedback en kritiek dat de stem van de betrokken individuen en gemeenschappen naar behoren kan worden gehoord en in acht kan worden genomen. Alleen door gewetensvol overleg met de belanghebbenden kan de ervaring die zij hebben opgedaan de basis vormen voor dit gezamenlijke streven naar de ontwikkeling van een duurzaam technologisch ecosysteem dat de bloei van de samenleving van morgen waarborgt.

09 BIJLAGEN

BIJLAGE 1: GLOSSARIUM

- **Verantwoording:** Verantwoording kan worden opgesplitst in twee subcomponenten: verantwoordingsplicht en controleerbaarheid. Verantwoordingsplicht verwijst naar de totstandbrenging van een ononderbroken keten van menselijke verantwoordelijkheid in de gehele workflow rondom AI-projecten. Het vereist ook dat verklaringen en rechtvaardigingen van zowel de inhoud van algoritmisch ondersteunde beslissingen als de processen achter de ontwikkeling ervan door bevoegde menselijke autoriteiten worden aangeboden in duidelijke, begrijpelijke en coherente taal. Controleerbaarheid ziet op de wijze waarop ontwikkelaars en gebruikers van AI-systemen verantwoordelijk kunnen worden gehouden. Dit aspect van verantwoording ziet op zowel de verantwoordelijkheid voor ontwikkeling en gebruik als voor de resultaten.

- **Algoritme:** Een algoritme is een procedure met instructie over hoe een reeks stappen genomen dient te worden om een bepaalde output te genereren. Een recept kan bijvoorbeeld worden beschouwd als een algoritme dat instructies geeft voor inputdata (i.e. de ingrediënten) en het creëren van een output (bv. een cake). In het geval van machinaal leren is het algoritme normaal gezien een reeks instructies die de software de opdracht geven met gebruik van een dataset (i.e. de input) een model te ontwikkelen of een onderliggend patroon te ontdekken (i.e. de output/het resultaat).

- **Algoritmische audits:** Er bestaan verschillende benaderingen van algoritmische audits, die variëren van de gerichte beoordeling van een systeem aan de hand van een zekere maatstaf (bv. mate van vooringenomenheid of 'bias') tot een bredere benadering die zich focust op de vraag of het systeem voldoet aan een reeks normen of aan een bepaalde regelgeving. Hoewel algoritmische audits meestal door professionals worden uitgevoerd met het oog op een onafhankelijke beoordeling, worden ze ook door journalisten, academici en activisten gebruikt als een middel om meer transparantie en verantwoording te verwezenlijken.

- **Geautomatiseerde beslissing:** Een geautomatiseerde beslissing is de selectie van een actie of een aanbeveling die wordt gemaakt met behulp van rekenkundige processen. Geautomatiseerde beslissingen verbeteren of vervangen het besluitvormingsproces dat gewoonlijk alleen door mensen wordt uitgevoerd. Meestal zijn geautomatiseerde beslissingen voorspellingen over personen of omstandigheden in de wereld die zijn afgeleid van een met machine learning uitgevoerde analyse van data over gebeurtenissen uit het verleden en de gelijkenis daarvan met een bepaalde reeks vooropgestelde omstandigheden.

- **Geautomatiseerd beslissysteem:** Een geautomatiseerd beslissysteem (ADS) ondersteunt of vervangt menselijke besluitvorming door gebruik te maken van rekenkundige processen om antwoorden op vragen te produceren, in de vorm van classificaties (bv. ja, nee; mannelijk, vrouwelijk, non-binaire; kwaadaardig, goedaardig) of scores (bv. mate van kredietwaardigheid, risico op het voorkomen van misdrijven, voorspelde tumorgroei). De meeste geautomatiseerde beslissystemen produceren voorspellingen over personen of omstandigheden met behulp van machinaal leren en andere rekenkundige logica door de waarschijnlijkheid te berekenen dat aan een bepaalde voorwaarde is voldaan.

Gewoonlijk wordt een geautomatiseerd beslissysteem "getraind" op historische data, waarbij wordt gezocht naar patronen tussen datapunten (bv. de relatie tussen de barometerstand, de omgevingstemperatuur en de sneeuwval). Een geautomatiseerde beslissing wordt genomen door bekende patronen te vergelijken met bestaande inputs om in te schatten hoe nauw ze bij elkaar aansluiten (bv. een weersvoorspelling op basis van de gelijkenis tussen de klimaatmetingen van vandaag en die uit het verleden). Voorbeelden van geautomatiseerde beslissystemen zijn algoritmen die kredietscores berekenen en biometrische herkenningssystemen die individuele personen trachten te identificeren op basis van fysieke kenmerken, zoals gelaatstrekken.

- **Vooringenomenheid door automatisering:** Vooringenomenheid door automatisering is een psychologisch verschijnsel dat zich kan voordoen wanneer de gebruikers van een AI-systeem de output van het systeem negeren of te nauw volgen, of niet in staat zijn de betrouwbaarheid van de beslissingen en uitkomsten van het systeem op passende wijze te beoordelen wegens technologische vooroordelen. Zo kan de gebruiker a) te veel op het systeem gaan vertrouwen, waardoor hij onnauwkeurige voorspellingen of classificaties niet opmerkt, of b) wantrouwig worden tegenover het systeem en het te weinig gebruiken, ondanks het feit dat het bij bepaalde taken beter kan presteren dan de gebruiker zelf.

- **Dataset:** Een dataset is een bestand met informatie dat bestaat uiteen verzameling metingen of waarnemingen, vastgelegd in een reeks rijen en kolommen. Elke rij komt overeen met een individu of een object dat kan worden beschreven aan de hand van een reeks geregistreerde waarden voor elk kenmerk dat door de reeks kolommen wordt weergegeven. Bijvoorbeeld, de volgende dataset vertegenwoordigt een reeks metingen voor patiënten in een fictieve dokterspraktijk, waarbij elke patiënt een uniek identificeerbaar patiëntnummer heeft:

Patiënt #	Leeftijd (bereik)	Gewicht (Kg)	Bloeddruk (mmHG)
1883652	26 > 30	71	115/75
1268833	31 > 40	nul	139/83
1776436	65 > 70	90	170/90
1557821	41 > 50	72	131/82

In het bovenstaande voorbeeld worden alleen de eerste 4 patiënten getoond, en worden slechts 3 kenmerken geregistreerd. Medische datasets kunnen echter enorm zijn, niet alleen wat betreft het aantal patiënten, maar ook wat betreft de mogelijke waarden die geregistreerd worden. Bovendien is er voor patiënt 1268833 geen registratie van zijn gewicht. Ontbrekende data vormen een belangrijke uitdaging voor machinaal leren, en kunnen de nauwkeurigheid van het ontwikkelde model beïnvloeden.

- **Procedurele gelijkheid:** Procedurele gelijkheid behelst het vereiste voor een persoon om een eerlijk proces te krijgen. Procedurele gelijkheid komt in de mensenrechten tot uiting in het recht op een behoorlijke verdediging, met inbegrip van het recht op juridische bijstand en het recht om getuigen op te roepen en aan een kruisverhoor te onderwerpen. Wanneer bij de strafvervolging gebruik wordt gemaakt van AI-technologieën, kan de procedurele gelijkheid inhouden dat de functies en de werking ervan kunnen worden geïnterpreteerd en betwist.

- **Uitlegbaarheid:** Nauw verbonden met transparantie, is de uitlegbaarheid van een AI-systeem, de mate is waarin de processen en de logica achter de resultaten van het systeem door menselijke gebruikers kunnen worden begrepen. Dit kan onder meer inhouden de mate waarin de innerlijke werking van het model in gewone taal kan worden omgezet, om betere besluitvorming en vertrouwen te bevorderen.

- **Billijkheid/Rechtvaardigheid:** Billijkheid kan op vele manieren worden gedefinieerd. Het kan worden uitgedrukt als de mate waarin een AI-systeem de invloed van bias in de input op de uitkomsten bevordert of voorkomt Omdat de levenscyclus van AI, met inbegrip van de beslissing om AI te gebruiken, in elke fase wordt beïnvloed door menselijke keuzes, wordt de billijkheid van AI bepaald door de menselijke vooringenomenheid en de invloed daarvan op wat AI doet en wie er wel of geen voordelen van ondervindt. In de context van AI vereist het waarborgen van billijkheid dat aandacht wordt besteed aan de gebruikte data, het algemene ontwerp van het systeem, de resultaten van het gebruik ervan, en de beslissingen over de implementatie ervan.
    - *Billijkheid van de data* houdt in dat de door AI gebruikte datasets voldoende representatief zijn voor de populatie, van hoge kwaliteit en relevantie zijn, dat de keuzes die initieel hebben geleid tot het verzamelen van de data worden onderzocht op vooringenomenheid, en dat de data controleerbaar zijn.
    - *Billijkheid bij het ontwerp* betekent dat de activiteiten van de systeemontwerpers doordacht en weloverwogen zijn, en rekening houden met de mogelijke vooringenomenheid van het ontwikkelingsteam. Billijkheid bij het ontwerpen vereist een evaluatie van de algemene probleemstelling en het gekozen resultaat, de selectie en het beheer van de gebruikte data, de selectie van kenmerken, en de vraag of voor leden van verschillende groepen en identiteiten vergelijkbare resultaten worden bereikt. Kortom, ontwerpers moeten ervoor zorgen dat de systemen die ze produceren niet bijdragen tot ongewenste sociale omstandigheden, zoals schadelijke discriminatie, uitputting van hulpbronnen, of onderdrukkende machtsstructuren.
    - *Billijkheid van de resultaten* is een beoordeling van de vraag of de beslissingen of andere resultaten die door AI worden geproduceerd rechtvaardig en eerlijk zijn, en resulteren in een rechtvaardige verdeling van rechten, plichten en publieke goederen. Billijkheid van de resultaten behelst ook een evaluatie van de waarden die door het gebruik van AI bevorderd of verhinderd worden.
    - We kunnen de billijkheid van een AI-systeem ook beoordelen op basis van de perspectieven van de belanghebbenden die het gebruik beïnvloeden of door het gebruik zelf beïnvloed worden. Elk AI-systeem heeft een andere en potentieel veranderende groep belanghebbenden. Algemene categorieën zijn subjecten, gebruikers, en samenlevingen.
    - Om de *billijkheid ten aanzien van de subjecten* vast te stellen, kunnen we ons afvragen of de persoon die het voorwerp is van een beslissing of actie die wordt genomen door of ondersteund door een AI-systeem, het proces en het resultaat als gerechtvaardigd en legitiem ervaart. Om de rechtvaardigheid en legitimiteit vast te stellen, moet de betrokkene wellicht weten hoe de beslissing of de actie tot stand is gekomen en welke factoren tot een ander resultaat hadden kunnen leiden (bv. een wervingsalgoritme wijst een sollicitant af, wat kan worden verklaard door aan te tonen dat de sollicitant een specifiek genoemde vaardigheid of kwalificatie ontbreekt). Een subject moet ook beroep kunnen aantekenen als hij het niet eens is met het resultaat (een sollicitant moet bijvoorbeeld aanvullende informatie kunnen vragen of de nauwkeurigheid van het wervingsalgoritme in twijfel kunnen trekken tegenover een mens met de bevoegdheid om het resultaat te wijzigen).

- De *billijkheid van de uitvoerders* kan uitgedrukt worden door middel van maatregelen om verantwoording af te leggen, zoals audit- en evaluatieprocessen. De uitvoerders hebben de taak ervoor te zorgen dat AI-systemen transparant en interpreteerbaar zijn voor diegenen die ze gebruiken en voor wie dat gebruik gevolgen heeft. Voorafgaand aan en tijdens het gebruik van AI moeten de uitvoerders rekening houden met de sociale, economische en politieke effecten, waarbij zij niet alleen oog moeten hebben voor de verwachte voordelen van AI, maar ook voor het optreden en het risico van schade en voor wie die draagt. Zo kan de invoering van een crimineel strafbepalingsalgoritme leiden tot meer consistentie in de rechtspraak en/of de besluitvorming stroomlijnen. Hetzelfde systeem kan echter ook discriminerende uitkomsten reproduceren, zoals wanneer gekleurde mensen langere straffen hebben gekregen voor vergelijkbare veroordelingen dan blanken in landen met een blanke meerderheid, als gevolg van een kenmerk in het ontwerp of de data die het gebruikt. Wanneer dergelijke problemen zich voordoen moet de functionele nauwkeurigheid of efficiëntie (indien aanwezig) van de AI-toepassing terzijde worden geschoven en moeten het ontwerp van het algoritme en model grondig worden geëvalueerd, met inbegrip van de beslissing of het al dan niet moet worden gebruikt.
- *Maatschappelijke rechtvaardigheid* is een breder punt van zorg. Een systeem waarvan het gebruik potentiële gevolgen heeft voor de rechten van individuen, groepen en/of de richting van de samenleving, vereist nauwlettende aandacht van de mens en een open discussie over het gebruik ervan. Beleidsmakers, academici en activisten hebben de taak om strategieën en acties voor te stellen en te bekritiseren die gericht zijn op het bevorderen van het algemeen welzijn en de sociale rechtvaardigheid. Wanneer AI wordt gebruikt in de private of de publieke sector (of in beide door publiek-private samenwerking), kan het bijdragen aan het in stand houden of juist aanvechten van bestaande sociale, economische en politieke verhoudingen.

  AI moet aldus onderworpen worden aan een open en inclusieve evaluatie van zijn rol in deze verhoudingen, en zij die betrokken zijn bij het ontwerp en de toepassing ervan moeten verantwoording afleggen voor hun keuzes. Uiteindelijk is het gebruik van AI, net als elk ander instrument, alleen aanvaardbaar indien dit betere levensomstandigheden kan creëren en geen schade veroorzaakt.

- **Generaliseerbaarheid:** Van een model wordt gezegd dat het generaliseerbaar is wanneer het doeltreffend is over een ruim spectrum van inputdata die de echte wereld weerspiegelen, en in een ruim spectrum van operationele contexten. Indien een model niet voldoende getraind is op representatieve data, zal het waarschijnlijk slechts in beperkte mate generaliseerbaar zijn wanneer het in de echte wereld wordt ingezet.

- **Intellectuele eigendom:** Intellectuele eigendom (IE) betreft de wettelijke bescherming van voortbrengselen van de menselijke geest, producten van creatief werk. De bekendste intellectuele eigendomsrechten zijn auteursrechten, octrooien, merken en bedrijfsgeheimen. Auteursrecht is een vorm van intellectuele eigendom die het recht van de maker beschermt om de vruchten te kunnen plukken van een origineel werk zoals een roman, muziekstuk of schilderij. Een octrooi is een exclusieve maar in de tijd beperkte licentie om voordeel te halen uit de uitvinding en ontdekking van nieuwe en nuttige werkwijzen, machines, productie-artikelen of samenstellingen van materie. Voorbeelden zijn nieuwe geneesmiddelen en technologieën voor zelfrijdende auto's. Met een merk kan een bedrijf zich het gebruik voorbehouden van een woord, naam, symbool of apparaat, of een combinatie daarvan, dat zijn goederen identificeert en onderscheidt van door anderen geproduceerde goederen. Een voorbeeld is de naam "Twitter" en de bijbehorende logo's die het bekende sociaal media-platform op unieke wijze identificeren en onderscheiden.

Een bedrijfsgeheim is elke informatie die gebruikt kan worden bij de exploitatie van een bedrijf of andere onderneming en die voldoende waardevol en geheim is om een feitelijk of potentieel economisch voordeel op te leveren ten opzichte van anderen, zoals het recept voor Coca-Cola.

- **Model:** Een model is het eindresultaat van de toepassing van een algoritme op een reeks inputdata (of variabelen) om een voorspellende of informatieve output-waarde te verkrijgen. Gewoonlijk is een model een formele (wiskundige) mappingfunctie waarmee wordt beoogd de onderliggende processen weer te geven, en de interacties daartussen, die worden verondersteld aanleiding te geven tot een verband tussen de waargenomen inputdata en de output van het algoritme. Het volgende eenvoudige model zou bijvoorbeeld de relatie kunnen weergeven tussen een reeks inputvariabelen, zoals de grootte van een onroerend goed (x1), het aantal slaapkamers (x2), de leeftijd van het onroerend goed (x3), en een outputvariabele (y), die de prijs weergeeft. Hier worden de coëfficiënten of parameters van de x-variabelen gebruikt als gewichten die aangeven hoe belangrijk elk van de inputvariabelen is, gebaseerd op de mate waarin ze y beïnvloeden. De taak van het lerende algoritme zou er in dit geval uit bestaan om de waarden voor elke parameter te vinden die de werkelijke huizenprijs in de trainingsdata nauwkeurig voorspellen. Het resulterende model zou dan gebruikt kunnen worden om de prijzen van nieuwe huizen te schatten, die niet in de oorspronkelijke dataset waren opgenomen.

- **Proportionaliteit:** Proportionaliteit is een rechtsbeginsel dat verwijst naar het idee dat een rechtvaardig resultaat moet worden bereikt op een manier die evenredig is met de kosten, de complexiteit en de beschikbare middelen. In dezelfde zin kan het ook worden gebruikt als een evaluatief begrip, zoals in het gegevensbeschermingsrecht, waar het inhoudt dat enkel persoonsgegevens mogen worden verzameld die noodzakelijk en toereikend zijn voor het doel van de verwerking.

- **Representativiteit:** De data die in het algoritme worden gebruikt, weerspiegelen de echte wereld. Is de gekozen steekproef representatief voor de kenmerken die in de algemene bevolking worden aangetroffen? Voorbeelden van niet-representativiteit vinden we bij de grootste beelddatabanken terug, die vaak worden samengesteld door mensen in een klein aantal landen. Een zoekopdracht naar "bruidsjurk" in een typische databank met afbeeldingen kan de huwelijkskledij van vele niet-westerse culturen niet altijd identificeren.

- **Sociaal-technisch systeem:** Een socio-technisch systeem is een systeem dat menselijk (of sociaal) gedrag koppelt aan het functioneren van een technisch systeem, en daarbij aanleiding geeft tot nieuwe (en opkomende) functies die noch tot de menselijke noch tot de technische elementen herleidbaar zijn. Door in te grijpen in menselijke gedragingen, attitudes, of hun relaties met de wereld, herstructureert het technische systeem menselijk gedrag. Het sociaal-technische perspectief is er een dat rekening houdt met de menselijke verlangens of doelen die een technologie beoogt te bereiken, of doet bereiken.

  AI-aanbevelingssystemen bijvoorbeeld, die vaak te vinden zijn op retail-, video- en sociale media websites, zijn sociaal-technisch omdat ze bedoeld zijn om door de beheerders van de site gewenst gedrag te produceren, zoals langere bezoektijden en/of de aankoop van goederen. Een algoritme voor machinaal leren op een site voor het delen van video's analyseert het kijkgedrag van duizenden of miljoenen gebruikers en doet aanbevelingen aan kijkers op basis van hun gelijkenis met een soortgelijke subset van gebruikers. Dit is een sociaal-technisch systeem, vermits het afhankelijk is van kennis over de kijkers en omdat het doel ervan is de aandacht van de kijkers voor de video's zo lang mogelijk vast te houden, wat reclame-inkomsten oplevert.

Als sociaal-technisch kunnen we ook die systemen beschrijven waarvan het bestaan, de uitvoering of de effecten gevolgen hebben voor de menselijke politieke, economische of sociale verhoudingen. Bewakingssystemen die door rechtshandhavingsinstanties gebruikt worden zijn bijvoorbeeld sociaal-technisch omdat hun invoering en gebruik politieke dimensies hebben; de geselecteerde doelwitten van politietoezicht worden acuter getroffen dan anderen door het gebruik van surveillance technologieën, gebaseerd op de historische keuzes die door overheden en rechtshandhavers zijn gemaakt. Vanuit dit sociaal-technische perspectief spelen surveillance-technologieën een rol in de relaties tussen mensen en de machtscentra in de samenleving.

- **Soft law:** Soft law verwijst naar beleids- en reguleringsinstrumenten die actie afdwingen of beperken zonder de kracht van overheidssancties of -straffen. Voorbeelden van soft law zijn "best practices" en ethische (of deontologische) richtsnoeren die door bedrijven en handelsverenigingen worden opgesteld. In sommige beroepen, zoals de advocatuur en de gezondheidszorg, is soft law het geheel van ethische praktijken die vereist zijn voor certificering. Schending van de medische ethiek kan leiden tot het verlies van de vergunning om de geneeskunde uit te oefenen. De bestraffende werking van deze regels op degenen die eraan onderworpen zijn, varieert. De Association of Computing Machinery (ACM) heeft bijvoorbeeld een "Code of Ethics and Professional Conduct" die door haar leden moet worden nageleefd. Er zijn echter geen voorgeschreven sancties en geen systeem van berechting voor leden van de vereniging die de code overtreden. Soft law kan ook ingezet worden om via positieve prikkels overheidsbeleid te realiseren. Denk aan belastingkredieten voor producenten van "groene" technologieën, waarmee bepaalde productiekeuzes worden gestimuleerd, maar niet afgedwongen.

- **Trainen/Testen van data:** Om een model te bouwen en ervoor te zorgen dat het accuraat is, zal een dataset doorgaans worden opgesplitst in twee kleinere sets: trainingsdata en testdata. De trainingsdata worden gebruikt om het model in eerste instantie te ontwikkelen, door de data in een algoritme in te voeren. Zodra het model is getraind, wordt het getest op de resterende data. Het doel van het opsplitsen van de data op deze manier is ervoor te zorgen dat het model kan generaliseren naar nieuwe omgevingen, aangezien de verzamelde data slechts een kleine steekproef van de totale populatie zullen vertegenwoordigen. Als alle data werden gebruikt om het model te trainen, bestaat het risico van *overfitting*, wat resulteert in een model dat goed presteert voor de oorspronkelijke dataset, maar slecht bij nieuwere data. Het testen van een model met "ongeziene" data stelt data-wetenschappers ook in staat om *underfitting* te identificeren, i.e. wanneer de mappingfunctie van een model niet strak genoeg op de verdeling van data past en daardoor niet in staat is om nauwkeurig de complexe patronen weer te geven die het probeert te classificeren of te voorspellen.

- **Transparantie:** De transparantie van AI-systemen kan betrekking hebben op verschillende kenmerken, zowel van hun innerlijke werking en gedragingen, als van de systemen en processen die hen ondersteunen. We kunnen stellen dat een AI-systeem transparant is als kan worden nagegaan hoe het is ontworpen, ontwikkeld en toegepast. Dit kan onder meer inhouden dat de data worden geregistreerd die zijn gebruikt om het systeem te trainen, of dat de parameters van het model dat de input (bv. een afbeelding) omzet in een output (bv. een beschrijving van de objecten in de afbeelding) werden bijgehouden. Het kan echter ook betrekking hebben op ruimere processen, zoals de vraag of er juridische belemmeringen zijn die individuën verhinderen toegang te verkrijgen tot informatie die nodig kan zijn om volledig te begrijpen hoe het systeem functioneert (bv. beperkingen in verband met intellectuele eigendom).

BIJLAGE 2: WERKZAAMHEDEN VAN DE RAAD VAN EUROPA EN ANDEREN OP HET VLAK VAN AI EN AANGRENZENDE GEBIEDEN TOT OP HEDEN

Dit aanvullende referentiemateriaal is geconsolideerd uit Hoofdstuk 4 van de *Haalbaarheidsstudie*. De nummering komt overeen met die in de *Haalbaarheidsstudie*.

4.1. Bescherming van persoonsgegevens

- Verdrag 108/108+ (1981/2018)
  - Verwerking van gevoelige gegevens kan alleen worden toegestaan indien passende richtsnoeren aanwezig zijn
  - Elk individu heeft het recht te weten wat het doel is van de verwerking van zijn gegevens. Daarnaast heeft hij recht op rectificatie en kennisneming wanneer gegevens worden verwerkt in strijd met de bepalingen van het verdrag
  - Transparantie, evenredigheid, verantwoording, effectbeoordelingen en respect voor privacy by design worden ingevoerd
  - Individuen mogen niet worden onderworpen aan besluiten die uitsluitend genomen worden op basis van geautomatiseerde gegevensverwerking zonder dat rekening wordt gehouden met persoonlijke standpunten
  - "Juridisch kader gebouwd rond het Verdrag blijft volledig van toepassing op AI-technologie, zodra de verwerkte gegevens binnen de reikwijdte van het Verdrag vallen."
  - Gemoderniseerd Verdrag 108+ aangenomen in 2018; Richtsnoeren inzake de bescherming van persoonsgegevens van kinderen in een onderwijscontext werden in november 2020 aangenomen
    - Stelt "de fundamentele beginselen van de kinderrechten in een onderwijscontext en bijstand aan wetgevers en beleidsmakers, verwerkingsverantwoordelijken en de industrie om deze rechten te handhaven" vast.

4.2. Cybercriminaliteit

- Verdrag inzake de bestrijding van strafbare feiten verbonden met elektronische netwerken ("Verdrag van Boedapest")(2001)
  - Omvat de "strafbaarstelling van inbreuken tegen en door middel van computers, alsook procedurele bevoegdheden om cybercriminaliteit te onderzoeken en elektronisch bewijs veilig te stellen".
    - Misdrijven omvatten, maar zijn niet beperkt tot, inbreuken op het auteursrecht, computergerelateerde fraude, kinderpornografie en inbreuken op een beveiligingsnetwerk
    - Het onderzoek omvat een reeks bevoegdheden en procedures, waaronder het onderscheppen en doorzoeken van computernetwerken
  - De hoofddoelstelling is "het voeren van een gemeenschappelijk strafrechtelijk beleid dat gericht is op de bescherming van de samenleving tegen cybercriminaliteit, in het bijzonder door middel van passende wetgeving en internationale samenwerking".
  - Het grensoverschrijdende karakter van digitale netwerken vereist een gecoördineerde internationale inspanning om misbruik van technologieën aan te pakken
  - Drie doelstellingen van het verdrag:
    - "Harmoniseren van de nationale strafrechtelijke elementen van inbreuken en de daarmee samenhangende bepalingen op het gebied van cybercriminaliteit."
    - "Het voorzien in binnenlandse strafprocesrechtelijke bevoegdheden die nodig zijn voor het onderzoeken en vervolgen van dergelijke inbreuken, alsmede van andere strafbare feiten die zijn gepleegd door middel van een computersysteem of waarvan het bewijsmateriaal zich in elektronische vorm bevindt."

        - "Opzetten van een snel en doeltreffend systeem voor internationale samenwerking."

4.3. Werkzaamheden op het gebied van algoritmische systemen

- Verklaring over de manipulatieve vermogens van algoritmische processen (2019)
    - Veel individuen zijn zich niet bewust van de gevaren van exploitatie van data
    - Computergestuurde middelen versterken bestaande vormen van discriminatie door individuen in categorieën te sorteren
    - Het Comité van Ministers vestigt de aandacht op "de toenemende dreiging die uitgaat van digitale technologieën voor het recht van mensen om onafhankelijk van geautomatiseerde systemen een mening te vormen en beslissingen te nemen."
    - De belangrijkste bedreigingen zijn micro-targeting, het identificeren van kwetsbaarheden en de herconfiguratie van sociale omgevingen
    - Het Comité geeft diverse aanbevelingen om deze bedreigingen aan te pakken, waaronder, maar niet beperkt tot, het overwegen van bijkomende beschermende maatregelen die zich concentreren op de gevolgen van gericht gebruik van technologieën, het initiëren van open, geïnformeerde en inclusieve publieke debatten over de grens tussen toelaatbare beïnvloeding en onaanvaardbare manipulatie, het mondiger maken van gebruikers door middel van een groter publiek bewustzijn en het bevorderen van digitale geletterdheid.
- Aanbeveling over de gevolgen van algoritmische systemen voor de mensenrechten (2020)
    - Lidstaten wordt geadviseerd hun wetgevende kaders, beleid en eigen praktijken te herzien om ervoor te zorgen dat de aanschaf, het ontwerp en de ontwikkeling van algoritmische systemen niet in strijd zijn met de mensenrechten
    - "Mensenrechten die vaak worden geschonden door de inzet van algoritmische systemen omvatten, maar zijn niet beperkt tot, het recht op een eerlijk proces; het recht op privacy en gegevensbescherming; het recht op vrijheid van gedachte, geweten en godsdienst; het recht op vrijheid van meningsuiting; het recht op vrijheid van vergadering; het recht op gelijke behandeling; en economische en sociale rechten."
    - Daarnaast wordt aanbevolen dat Lidstaten regelmatig, inclusief en transparant overleg plegen met de relevante belanghebbenden, waarbij de nadruk ligt op de stem van kwetsbare groepen.
    - Deze aanbeveling omvat diverse verplichtingen voor staten met betrekking tot de bescherming en bevordering van mensenrechten en fundamentele vrijheden in de context van algoritmische systemen, onder meer op het vlak van wetgeving, transparantie, verantwoording en doeltreffende rechtsmiddelen, voorzorgsmaatregelen, enz.
- MSI-AUT Verantwoordelijkheid en AI: Een onderzoek naar de implicaties van geavanceerde digitale technologieën (waaronder AI-systemen) voor het concept van verantwoordelijkheid binnen een mensenrechtenkader (2019)
    - Dit rapport schetst wat AI is en hoe taakspecifieke technologieën werken, en schetst bedreigingen en schade die gepaard gaan met geavanceerde digitale technologieën, en een reeks "verantwoordelijkheidsmodellen" voor de nadelige gevolgen van AI-systemen
    - De belangrijkste aanbevelingen uit dit rapport zijn "effectieve en legitieme mechanismen die mensenrechtenschendingen zullen verhinderen en voorkomen", beleidskeuzes met betrekking tot verantwoordelijkheidsmodellen voor AI-systemen, ondersteuning van technisch onderzoek met betrekking tot de bescherming van mensenrechten en "algoritmische audits", en de aanwezigheid van legitieme bestuursmechanismen voor de bescherming van mensenrechten in het digitale tijdperk
    - Diegenen die digitale technologieën ontwikkelen en implementeren, kunnen dat niet doen zonder verantwoordelijkheid - zij moeten ter verantwoording worden geroepen voor nadelige gevolgen

4.4. Werkzaamheden op het gebied van justitie

- Europees Ethisch Handvest over het gebruik van AI in gerechtelijke systemen en hun omgeving (2018)
  - In dit handvest worden vijf grondbeginselen uiteengezet, waaronder respect voor fundamentele rechten, non-discriminatie, kwaliteit en veiligheid, transparantie, onpartijdigheid en rechtvaardigheid, en "onder controle van de gebruiker".
  - De meeste toepassingen van AI op justitieel gebied blijken in de private sector te liggen - "commerciële initiatieven gericht op verzekeringsmaatschappijen, juridische diensten, advocaten en particulieren."
  - Enkele mogelijke toepassingen van AI in een gerechtelijke context zijn verbetering van de jurisprudentie, toegang tot het recht en het creëren van nieuwe strategische instrumenten
  - Andere toepassingen die aanzienlijke methodologische voorzorgsmaatregelen vereisen, omvatten het opstellen van schalen, de ondersteuning van alternatieve maatregelen voor geschillenbeslechting in burgerlijke zaken, online geschillenbeslechting in de precontentieuze fase (wanneer een later beroep bij de rechter mogelijk blijft), of de identificatie van de plaats waar strafbare feiten worden gepleegd

4.5. Werkzaamheden op het gebied van goed bestuur en verkiezingen

- Europees Comité inzake Democratie en Governance (CDDG)
  - Werkt momenteel aan een studie over de impact van digitale transformatie op democratie en governance
- Commissie van Venetië: Beginselen voor een grondrechtenconform gebruik van digitale technologieën in verkiezingsprocedures (2020)
  - Benadrukt acht beginselen voor het gebruik van digitale technologieën bij verkiezingen in overeenstemming met de mensenrechten
  - Deze acht beginselen (in het document in detail beschreven) zijn:
    - 1. "De beginselen van vrijheid van meningsuiting die een robuust publiek debat impliceren, moeten worden vertaald naar de digitale omgeving, in het bijzonder tijdens verkiezingen."
    - 2. "Tijdens verkiezingen moet een bevoegd onpartijdig orgaan zoals een kiesraad of een gerechtelijk orgaan de bevoegdheid krijgen om private bedrijven te verplichten duidelijk omschreven inhoud van derden van het internet te verwijderen - op basis van kieswetten en in overeenstemming met internationale normen."
    - 3. "Tijdens verkiezingen moeten het open internet en de netneutraliteit worden beschermd."
    - 4. "Persoonsgegevens moeten doeltreffend worden beschermd, in het bijzonder tijdens de cruciale periode van verkiezingen."
    - 5. "De integriteit van verkiezingen moet worden bewaard door middel van periodiek herziene regels en voorschriften inzake politieke reclame en de verantwoordelijkheid van tussenpersonen op het internet."
    - 6. "De integriteit van verkiezingen moet worden gewaarborgd door de specifieke internationale regelgeving aan te passen aan de nieuwe technologische context en door institutionele capaciteiten te ontwikkelen om cyberdreigingen te bestrijden."
    - 7. "Het internationale samenwerkingskader en de publiek-private samenwerking moeten worden versterkt ."
    - 8. "De invoering van zelfregulering moet worden bevorderd."

4.6. Werkzaamheden op het gebied van gendergelijkheid en non-discriminatie

- Aanbeveling CM/Rec(2019)1 van het Comité van Ministers betreffende het voorkomen en bestrijden van seksisme
  - De aanbeveling stelt dat er maatregelen moeten worden genomen om seksisme te voorkomen en te bestrijden, en bevat tevens een oproep om een perspectief van gendergelijkheid te integreren in alle werkzaamheden in verband met AI, en tegelijkertijd manieren te vinden om de genderkloof en seksisme te helpen elimineren
- Europese Commissie tegen racisme en intolerantie (ECRI) - Discriminatie, artificiële intelligentie en algoritmische besluitvorming (2018)
  - AI-toepassingen weten "aan de huidige wetten te ontsnappen". Het merendeel van de wetgeving omtrent non-discriminatie heeft alleen betrekking op specifieke beschermde kenmerken. Er zijn andere vormen van discriminatie die niet gecorreleerd zijn met beschermde kenmerken, maar die toch sociale ongelijkheid kunnen versterken
  - Het idee van sectorspecifieke regels voor de bescherming van billijkheid en mensenrechten op het gebied van AI wordt voorgesteld, aangezien verschillende sectoren verschillende waarden en problemen vereisen
  - Voor een bepaalde sector formuleert ECRI verschillende vragen die beantwoord moeten worden:
    - "Welke regels zijn van toepassing in deze sector, en wat zijn de achterliggende beweegredenen?"
    - "Hoe wordt of kan AI-besluitvorming in deze sector gebruikt worden, en wat zijn de risico's?"
    - "Moet de wet worden verbeterd in het licht van AI-besluitvorming, gelet op de beweegredenen voor de regels in deze sector?"

4.7. Werkzaamheden op het gebied van onderwijs en cultuur

- Aanbeveling CM/Rec(2019)10 van het Comité van Ministers over de ontwikkeling en bevordering van digitaal burgerschapsonderwijs
  - Nodigt Lidstaten uit regelgevende beleidsmaatregelen inzake digitaal burgerschapsonderwijs vast te stellen, alle relevante belanghebbenden te betrekken bij het ontwerp, de uitvoering en de evaluatie van wetgeving, beleid en praktijken op het gebied digitaal burgerschapsonderwijs, en de doeltreffendheid van het nieuwe beleid en nieuwe praktijken te evalueren
  - Benadrukt het belang van "het in staat stellen van burgers om de vaardigheden en competenties te verwerven voor een democratische cultuur, door hen in staat te stellen de uitdagingen en risico's aan te gaan die voortvloeien uit de digitale omgeving en opkomende technologieën".
- Stuurgroep voor onderwijsbeleid en -praktijk (CDPPE)
  - Verkent de implicaties van het gebruik van AI in onderwijsomgevingen
- Eurimages en de Raad van Europa - “Entering the new paradigm of artificial intelligence and series” (2019)
  - Studie over de impact van voorspellende technologieën en AI op de audiovisuele sector
  - In dit document wordt het gebruik van artificiële intelligentie in de audiovisuele sector aangemerkt als "een potentieel gevaar voor de diversiteit van de inhoud en de vrije toegang tot informatie van de burgers van de lidstaten".
  - Er worden vijf slot-aanbevelingen gedaan, gaande van "Eurimages een mandaat geven om bekwaamheid op het gebied van Series op te bouwen", tot "handelsvoorwaarden voor de productie van series in de lidstaten voorstellen die geïnspireerd zijn op internationale goede praktijken en samenwerking aanmoedigen" en "het publiek bewustzijn van de gevolgen van AI in de audiovisuele sector vergroten".

- Een andere aanbeveling is dat de Raad van Europa de oprichting zou overwegen van een "bestuursorgaan voor een AI-certificering voor de media".

4.8. Werkzaamheden van de Parlementaire Vergadering van de Raad van Europa

- Technologische convergentie, artificiële intelligentie en mensenrechten (2017)
  - Roept op tot een implementatie van "een echte wereldwijde internet governance die niet afhankelijk is van private belangengroepen of slechts een handvol staten."
  - Daarnaast roept de Vergadering het Comité van Ministers op om:
    - "de modernisering van het Verdrag tot bescherming van personen met betrekking tot de geautomatiseerde verwerking van persoonsgegevens af te ronden"
    - "een kader vast te stellen voor zowel ondersteunende technologieën als zorgrobots in de Strategie 2017-2023 van de Raad van Europa inzake inclusie van personen met een handicap ".
  - De Vergadering wijst ook nogmaals op het belang van verantwoording en verantwoordelijkheid van AI-systemen die bij mensen geplaatst worden, het informeren van het publiek over het genereren en verwerken van hun persoonsgegevens, en het erkennen van rechten in verband met de eerbiediging van het privé- en gezinsleven, naast andere voorgestelde richtsnoeren
- 7 rapporten over AI zijn door de Parlementaire Vergadering aangenomen met onderwerpen variërend van democratisch bestuur tot discriminatie en de juridische aspecten van autonome voertuigen
- Noodzaak van democratisch bestuur van artificiële intelligentie (2020)
  - De Vergadering beveelt het volgende aan:
    - "De uitwerking van een juridisch bindend instrument dat artificiële intelligentie regelt..."
    - "Ervoor zorgen dat een dergelijk juridisch bindend instrument gebaseerd is op een alomvattende aanpak, de hele levenscyclus van AI-systemen omvat, gericht is tot alle belanghebbenden, en mechanismen bevat om de tenuitvoerlegging van dit instrument te waarborgen."

4.9. Werkzaamheden van het Congres van Lokale en Regionale Overheden van de Raad van Europa

- De voorbereiding van "Smart cities: de uitdagingen voor de democratie" is aan de gang en zal in de tweede helft van 2021 worden uitgebracht

4.10. Werkzaamheden van de Commissaris voor de Rechten van de Mens

- Artificiële intelligentie uitpakken: 10 stappen ter bescherming van de mensenrechten (2019)
  - Formuleert aanbevelingen om negatieve gevolgen van AI-systemen voor de mensenrechten te verzachten of te voorkomen
  - Geeft praktische aanbevelingen met 10 actiegebieden: mensenrechten effectbeoordelingen; openbare raadplegingen; normen omtrent mensenrechten in de private sector; informatie en transparantie; onafhankelijk toezicht; non-discriminatie en gelijkheid; gegevensbescherming en privacy; vrijheid van meningsuiting, vrijheid van vergadering en vereniging, en het recht op werk; verhaalsmogelijkheden; en het bevorderen van kennis en begrip van AI
  - Er wordt een checklist beschikbaar gemaakt om de aanbevelingen in het document te kunnen operationaliseren

4.11. Werkzaamheden van de Raad van Europa op het gebied van jeugdzaken

- Raad van Europa Strategie voor de Jeugdsector 2030 (2020)
  - Roept op tot een verbetering van de institutionele antwoorden op nieuwe kwesties (waaronder AI) die van invloed zijn op de rechten van jongeren en hun overgang naar volwassenheid
  - De drie belangrijkste aandachtspunten van de strategie voor 2030 zijn:
    - "Het verbreden van de participatie van jongeren."
    - "De toegang van jongeren tot rechten versterken."
    - "De kennis van jongeren verdiepen.
  - Andere thematische prioriteiten zijn onder meer het vergroten van de capaciteit voor participatieve democratie, beleidsvoering met betrokkenheid van diverse groepen jongeren, het versterken van "de capaciteiten, de daadkracht en het leiderschap van jongeren om geweld te voorkomen, conflicten om te buigen en een cultuur van vrede op te bouwen...", naast verschillende andere prioriteiten

4.12. Werkzaamheden van de Europees Commissie voor strafrechtelijke vraagstukken (CDPC)

- Haalbaarheidsstudie over een toekomstig instrument van de Raad van Europa inzake artificiële intelligentie en strafrecht (2020)
  - Een werkgroep van het CDPC heeft in december 2019 de opdracht gekregen "een haalbaarheidsstudie uit te voeren waarin de reikwijdte en de belangrijkste elementen van een toekomstig instrument van de Raad van Europa over AI en strafrecht, bij voorkeur een verdrag, worden vastgesteld"
  - Onderzoekt de mogelijkheden van de Raad van Europa om het pad te effenen voor een internationaal juridisch instrument inzake AI en strafrecht en stelt, op basis van de antwoorden van de lidstaten op vragenlijsten over AI en strafrecht, de belangrijkste elementen vast van een internationaal instrument van de Raad van Europa inzake AI en strafrecht
  - Identificeert vier doelstellingen van het rechtsinstrument:
    i. Het tot stand brengen van een internationaal kader voor de ontwikkeling van nationale wetgeving inzake strafrechtelijke vraagstukken in verband met AI (in het bijzonder in de context van automatisering van voertuigen);
    ii. Lidstaten aanmoedigen juridische uitdagingen op het gebied van strafrecht en AI aan te pakken via wetgeving, gesteund op gemeenschappelijke normatieve beginselen;
    iii. Anticiperen op reeds geconstateerde bewijsproblemen en andere juridische problemen in verband met strafrechtelijke aansprakelijkheid en AI, en zorgen voor eerlijke procesbeginselen en doeltreffende internationale samenwerking op dit gebied; en
    iv. Garanderen dat AI-systemen worden ontwikkeld in overeenstemming met de fundamentele rechten die door de instrumenten van de Raad van Europa worden beschermd.
  - Conclusie van de studie: "het bereiken van overeenstemming over gemeenschappelijke normen voor een duidelijke en correcte toewijzing van mogelijke strafrechtelijke verantwoordelijkheid en voor het verduidelijken van daarmee verband houdende procedurele kwesties en mogelijke implicaties voor de mensenrechten, moet een gezamenlijke inspanning zijn van actoren uit de publieke en de private sector, zodat de technologie zich met succes kan ontwikkelen op een wijze die de grondbeginselen van onze burgerlijke samenleving eerbiedigt."