

Artificiële Intelligentie

Human-in-Command

Voorwoord

AI heeft het potentieel om aanzienlijke voordelen voor de samenleving te hebben, maar AI brengt ook risico's en complexe uitdagingen met zich mee.

Artificiële Intelligentie (AI) maakt al enige tijd een groot aantal belangrijke ontwikkelingen door en heeft inmiddels zeer veel toepassingen in de maatschappij. AI-systemen hebben daarmee een steeds grotere en diepere impact op ons dagelijks leven.

En hoewel AI grote maatschappelijke voordelen zou kunnen hebben, brengt het ook risico's en complexe beleidsuitdagingen met zich mee op gebieden als veiligheid en controleerbaarheid, social-economische aspecten, ethiek, betrouwbaarheid, wet- en regelgeving, etc. Dat vraagt om kritische reflectie en politieke keuzes maar ook om samenwerking tussen makers, gebruikers en de maatschappij als geheel.

ALLAI is ervan overtuigd dat de voordelen van AI alleen duurzaam kunnen worden bereikt als ook de uitdagingen rondom AI adequaat en tijdig worden geadresseerd.

Dit position paper geeft handvatten voor verantwoord(e) ontwikkeling, inzet en gebruik van AI.

mr. Catelijne Muller
Voorzitter ALLAI

RAI

Human-in-Command

Technologie overkomt ons niet. Ook AI niet. Het is aan ons om te bepalen of, wanneer en hoe we AI gebruiken.

ALLAI roept al enige tijd op tot een 'human-in-command' benadering van AI waarbij machines machines blijven en mensen te allen tijde de controle over deze machines behouden.

Maar human-in-command verwijst ook naar de capaciteit om de algehele activiteit van het AI-systeem te overzien (inclusief zijn bredere economische, maatschappelijke, juridische en ethische gevolgen).

En het verwijst naar het vermogen om te kiezen wanneer en hoe het systeem in iedere specifieke situatie wordt gebruikt.

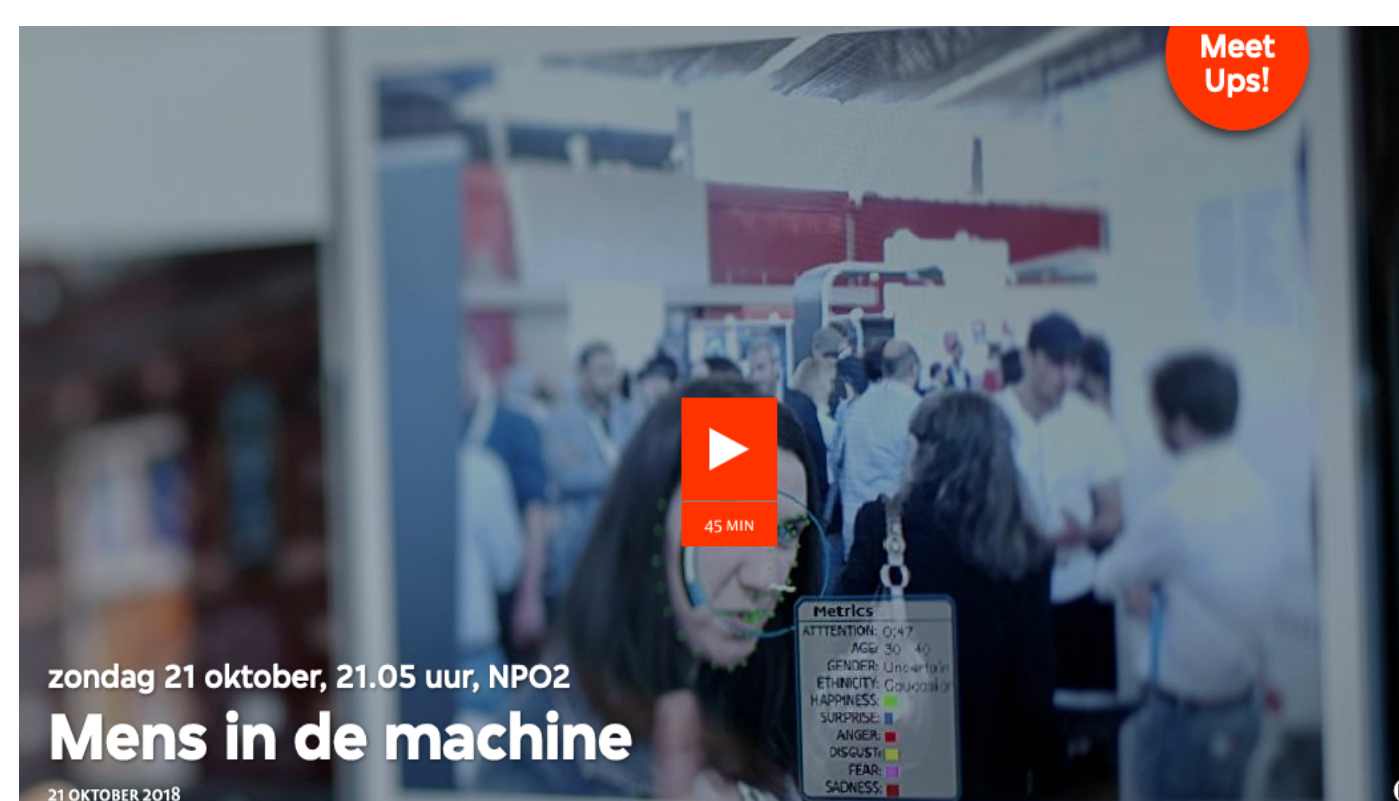
Daarbij kan het bijvoorbeeld gaan om de keuze om een AI-systeem in een bepaalde situatie niet te gebruiken, om een bepaalde mate van menselijk oordeel vast te stellen tijdens het gebruik van het systeem of om te garanderen dat een door het systeem genomen beslissing kan worden herroepen.

Europese Commissie

De human-in-command benadering is inmiddels overgenomen door de EU High Level Expert Group on AI (AI HLEG) en de Europese Commissie.

Een van de elementen om te kunnen voldoen aan het eerste vereiste voor Betrouwbare AI (Trustworthy AI, zoals omschreven in de Ethische Richtlijnen voor Betrouwbare AI van de AI HELG) is 'menselijke autonomie en toezicht', ondersteund door het 'human-in-command-principe'.

VPRO Tegenlicht: Mens in de Machine



HUMAN
IN COMMAND

Nederland, Europa en de wereld

Europese Unie

Met name op EU niveau, maar ook internationaal en in een groot aantal EU-lidstaten staat AI allang op de agenda.

In 2017 signaleerde het EESC (Europees Economisch en Sociaal Comité) in haar baanbrekend rapport *AI & Society*, dat AI een technologie was om serieus rekening mee te houden.* Het rapport benoemde de belangrijkste impact domeinen van AI en riep de Europese Commissie en maatschappelijke spelers op om gezamenlijk de (rand)voorwaarden voor de verdere verantwoorde ontwikkeling en inzet van AI te bepalen en hierin mondiaal een voortrekkersrol te spelen.

Deze oproep heeft op EU-niveau ruimschoots zijn gehoor gekregen in de vorm van:

- Een Europese AI strategie, *AI for Europe*
- De *High Level Expert Group on AI* van de Europese Commissie
- Ethische Richtlijnen voor Betrouwbare AI van de AI HLEG
- Investerings- en Beleidsaanbevelingen voor Betrouwbare AI van de AI HLEG
- Update van de Assessment List uit de Ethics Guidelines for Trustworthy AI
- Whitepaper Artificial Intelligence Europese Commissie (19 februari 2020)
- De EU zoekt nadrukkelijk aansluiting met derde-landen zoals Canada en Japan

Raad van Europa

De Raad van Europa is een proces gestart om de haalbaarheid van een juridisch bindend instrument voor AI op verdragsniveau te onderzoeken. ALLAI voorzitter Cateijne Muller is door de Raad van Europa aangetrokken als expert.

Nederland

Nederland heeft op 8 oktober jl. haar AI strategie gepresenteerd. Het AI-beleid dat daarin wordt omschreven, is echter zeer gefragmenteerd en versnipperd over diverse ministeries, waardoor het risico bestaat dat het onvoldoende samenhangt, niet beklijft of zelfs vervliegt.

De strategie biedt bovendien onvoldoende visie op de grote kansen maar ook uitdagingen die AI biedt. Naar de mening van ALLAI is het Nederlandse kabinet op beide punten te 'nonchalant'. De kansen worden onvoldoende ondersteund met financiële middelen en de uitdagingen worden weliswaar benoemd, maar niet daadwerkelijk onderkend of geadresseerd.

Exemplarisch voor dat laatste is de recente SyRI-uitspraak, waarin de Rechtbank heeft bepaald dat de SyRI-wet, die een juridische grond biedt om met toepassing van AI risicovoorspellingen te doen ten aanzien van fraude met overheidsregelingen, in strijd is met art. 8 EVRM.

*EESC INT/806: Rapporteur Cateijne Muller (voorzitter ALLAI), expert Robert Went (WRR)


```
self.file = None
self.fingerprints = set()
self.logdups = True
self.debug = debug
self.logger = logging.getLogger(__name__)
if path:
    self.file = open(os.path.join(path, 'fingerprint.log'), 'a')
    self.file.seek(0)
    self.fingerprints.update(e.request for e in self.logdups)
```

```
classmethod
def from_settings(cls, settings):
    debug = settings.getbool('SUPERFILTER_DEBUG')
    return cls(job_dir(settings), debug)
```

```
def request_seen(self, request):
    fp = self.request_fingerprint(request)
    if fp in self.fingerprints:
        return True
    self.fingerprints.add(fp)
    if self.file:
        self.file.write(fp + os.linesep)
```

```
def request_fingerprint(self, request):
    return request_fingerprint(request)
```


Wat is Artificiële Intelligentie?

Het doel van AI onderzoek en ontwikkeling is de automatisering van intelligent gedrag.

Containerbegrip

AI is een containerbegrip voor een groot aantal (sub)domeinen zoals *machine learning*, *deep learning*, *natural language processing*, etc.

Narrow versus General AI

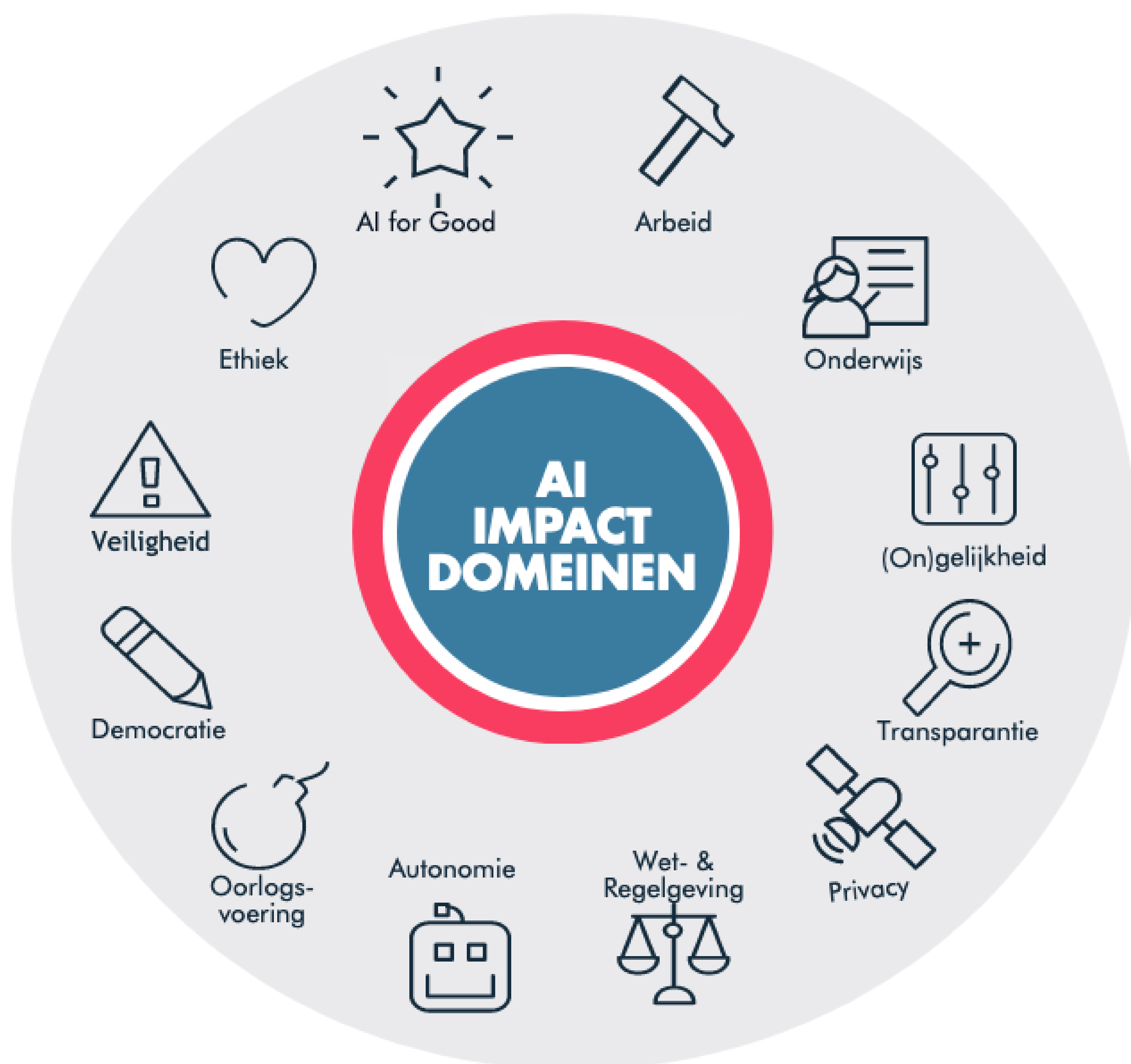
AI valt grofweg te onderscheiden in *narrow AI* (in staat om zeer specifieke taken te verrichten) en *general AI* (in staat iedere intellectuele taak te verrichten die een mens kan verrichten). *General AI* bestaat niet.

Hoewel er bedrijven en onderzoekinstellingen zijn die zich volledig richten op *General AI*, zijn wetenschappers het er niet over eens of en zo ja, wanneer *general AI* zal worden bereikt.

De huidige AI-systemen zijn voorbeelden van *narrow AI* met belangrijke karakteristieken die hiernaast staan omschreven.

Karakteristieken AI

- Doelgericht: AI-systemen definiëren niet hun eigen doelen, maar gebruiken technieken om de optimale oplossing voor een door mensen gesteld doel te bereiken.
- Big Data: AI-systemen hebben veel data nodig om goed te presteren en de kwaliteit, het volume en de inhoud van de data beïnvloedt de werking van het AI-systeem.
- Correlatie: AI-besluiten berusten veelal slechts op statistische verbanden opgesteld uit veel data (correlatie). Het systeem legt (of ziet) geen causaal verband tussen een gebeurtenis of handeling en een beslissing, maar doet slechts een voorspelling.
- Black-boxes: veel AI-systemen zijn zogenaamde black boxes, waarbinnen (beslis)processen plaatsvinden die niet kunnen worden doorgrond.
- Gezond verstand: AI-systemen hebben geen gezond verstand.
- Algoritmes zijn slechts een onderdeel van AI: een AI-systeem bestaat uit (vaak meerdere) complexe algoritmes, data en modellen.



AI Impact Domeinen

De impact van AI, zowel positief als negatief, is zeer breed en raakt aan economische, ethische, sociaal-maatschappelijke en juridische vraagstukken.

Veiligheid

Om veilig te kunnen opereren moeten AI-systemen robuust zijn, bestand zijn tegen hacken en veilig opereren onder onbekende, kritieke of onvoorspelbare situaties.

AI-systemen mogen alleen in gebruik mogen worden genomen indien deze voldoen aan specifieke eisen voor veiligheid. Deze eisen dienen door beleidsmakers, AI- en veiligheidsspecialisten, bedrijven en maatschappelijke organisaties gezamenlijk te worden bepaald.

Ethiek

AI roept veel ethische vragen op. Wat is de invloed van autonome (zelflerende) AI op onze persoonlijke integriteit, autonomie, waardigheid, zelfstandigheid, gelijkheid, veiligheid, keuzevrijheid? Hoe zorgen we ervoor dat onze fundamentele normen, waarden en mensenrechten geëerbiedigd en gewaarborgd blijven?

AI-systemen bevatten geen ethische waarden. Ook de inzet en het gebruik ervan vindt regelmatig plaats voor ongewenste doeleinden of met onvoldoende oog voor de sociaal-maatschappelijke of ethische gevolgen.

In de eerste plaats zijn er de trainingsdata. Deze data dienen juist te zijn, van goede kwaliteit, gevarieerd, voldoende diepgaand en onbevooroordeeld. Ten tweede vindt de ontwikkeling van AI op dit moment vaak binnen een homogene omgeving van voornamelijk jonge, blanke mannen plaats, waardoor (al dan niet bewust) cultuur- en genderdispariteiten in AI worden ingebed.

Dit betekent dat niet alleen de ontwikkeling, maar ook de inzet en het gebruik van AI-systemen (zowel publiek als commercieel) dient plaats te vinden binnen de grenzen van onze grondrechten, wet- en regelgeving en ethische normen en waarden.

De AI HLEG heeft inmiddels Ethische Richtlijnen voor Verantwoorde AI gepresenteerd, waarbij de grondrechten als basis voor AI-ethiek zijn genomen.

De richtlijnen beschrijven 7 vereisten voor betrouwbare (ethische, wettige en robuuste) AI:

- Menselijke controle en menselijk toezicht
- Technische robuustheid en veiligheid
- Privacy en datagovernance
- Transparantie
- Diversiteit, non-discriminatie en rechtvaardigheid
- Maatschappelijk en milieuwelzijn
- Verantwoording

Begrijpelijkheid, controleerbaarheid en uitlegbaarheid

Acties en beslissingen van AI-systemen grijpen in toenemende mate in in het leven van mensen. Voorbeelden zijn het gebruik van AI bij *predictive policing*, bij de beoordeling van hypotheekaanvragen of bij het toelatingsproces tot verzekeringen.

Acceptatie en duurzame ontwikkeling en toepassing van AI hangt samen met de mogelijkheid om de werking, acties en beslissingen van AI-systemen te begrijpen, controleren en verklaren, met name ook achteraf.

Inmiddels wordt er alom gepleit voor transparante, begrijpelijke, uitlegbare en controleerbare AI-systemen waarvan de werking ook achteraf verklaarbaar is. Dit is cruciaal en vaak zelfs verplicht, maar niet voldoende. Belangrijk is om ook vast te stellen welke beslisprocessen of -momenten aan AI-systemen kunnen worden overgelaten en welke niet en wanneer menselijke interventie gewenst of vereist zou moeten zijn (human-in-command).

Werk

AI kan grote voordelen hebben wanneer het wordt ingezet bij gevaarlijk, zwaar, uitputtend, vuil, onaangenaam, repetitief of saai werk, maar ook dataverwerking en -analyse, planning en predictie, kan in toenemende mate door AI-systemen worden verricht.

AI zal niet alleen gevolgen hebben voor de hoeveelheid beschikbaar werk, maar ook voor de aard van het bestaande werk en de verhoudingen tussen werkgever en werknemer. AI-systemen worden steeds meer ingezet voor het volgen en monitoren van werkers, waardoor de autonomie en privacy in het geding komen. Werk wordt meer en meer bepaald en verdeeld door algoritmes, zonder menselijke tussenkomst, wat de aard van het werk en de werkomstandigheden beïnvloedt.

Er bestaat het risico op verschraling van de werkzaamheden en het verleren van belangrijke vaardigheden bij gebruik van AI-systemen.

Overheid en sociale partners hebben de mogelijkheid om de manier waarop AI verder wordt ontwikkeld en wordt toegepast in het arbeidsproces te bepalen en moeten deze mogelijkheid ook met beide handen aanpakken.

Het is daarbij belangrijk om niet alleen te focussen op waar AI toe in staat is, maar juist ook op waar mensen toe in staat zijn (creativiteit, empathie, samenwerken), op wat we willen dat mensen blijven doen en om te zoeken naar mogelijkheden om mens en machine beter samen te laten werken (complementariteit).

Augmented intelligence, waarbij mens en machine samenwerken en elkaar versterken, is de interessantste toepassing van AI omdat het gaat om mens mét machine, in tegenstelling tot mens in plaats van machine.

Educatie en Skills

Het bijhouden of aanleren van digitale vaardigheden is nodig om mensen de mogelijkheid te geven om zich aan de snelle ontwikkelingen op het gebied van AI aan te passen.

Niet iedereen zal echter in staat of geïnteresseerd zijn om te coderen en programmeur te worden. Beleid en financiële middelen zullen daarom ook gericht moeten worden op educatie en ontwikkeling van vaardigheden op terreinen die niet door AI-systemen zullen worden bedreigd.

Wanneer wordt ingezet op complementariteit van mens en AI (*augmented intelligence*) zal voor iedereen en vanaf jonge leeftijd educatie in de omgang en het werken met AI-systemen noodzakelijk zijn om te zorgen dat mensen autonomie en controle kunnen behouden in hun leven en werk.

THE GREAT HACK

De recente Netflix documentaire "The Great Hack" geeft inzicht in de wijze waarop Cambridge Analytica met behulp van grote hoeveelheden data en AI in staat was kiezers te profileren en deze te proberen met fake news te beïnvloeden (nudgen).

Bedrijfsmacht

Het overgrote deel van de ontwikkeling van AI en alle daarbij behorende elementen (ontwikkelpplatforms, data, kennis, expertise) is in handen van de 5 grote Amerikaanse en 2 Chinese technologiebedrijven (Amazon, Apple, Facebook, Google, Microsoft, Alibaba, Tencent).

ALLAI beveelt aan om bij de beoordeling van marktmacht rekening te houden met het volume van data - de bouwsteen van veel AI-systemen - voor het toepassen van regels inzake concurrentiebeperkend gedrag, misbruik van dominantie en bij het evalueren van fusies.

Inclusiviteit

De EU, (inter)nationale beleidsmakers en maatschappelijke organisaties hebben daarnaast een belangrijke rol te vervullen door ervoor te zorgen dat AI-systemen voor iedereen toegankelijk zijn, maar ook dat de ontwikkeling ervan in een open omgeving plaatsvindt.

Democratie

Het gericht gebruik van AI-systemen voor bijvoorbeeld nieuwsaggregatie op bijvoorbeeld social media, kan leiden tot vernauwing van informatie en splijting van de maatschappij (denk aan 'filter bubbles' en 'nepnieuws' op Twitter en Facebook tijdens de Amerikaanse verkiezingen en de Brexit stemming).

AI is in staat om, door middel van analyse van zeer veel (vaak persoonlijke) data, voorspellingen over gedrag en voorkeuren te maken, en zo de keuzes van mensen en organisaties (van commerciële beslissingen tot verkiezingen en referenda) te bepalen of beïnvloeden.

Dit is een bedreiging voor een eerlijke en open democratie. In het huidige tijdperk van polarisering en ontmanteling van internationale instituties, kan de precisie en kracht van dergelijke propaganda-technologie snel voor verder ontwrichting van de maatschappij zorgen.

Oorlogsvoering

Momenteel houdt een onbekend aantal landen en sectoren zich bezig met onderzoek naar en ontwikkeling van dodelijke autonome wapensystemen. Dit brengt grote punten van zorg met zich mee, zoals het risico op een oncontroleerbare wapenwedloop van historisch ongekennde omvang, militaire situaties waarbij er vrijwel geen sprake meer is van menselijke controle, 'oneindige oorlogen' en het risico op catastrofaal systeemfalen.

Het Europees Parlement heeft opgeroepen tot de urgente ontwikkeling van een gemeenschappelijk, juridisch bindend standpunt waarin wordt ingegaan op ethische en juridische vragen op het gebied van menselijke controle, toezicht, verantwoording en de uitvoering van internationale mensenrechtenwetgeving, internationale humanitaire wetgeving en militaire strategieën. De HLEG AI staat achter de resolutie en alle verwante inspanningen. Ook ALLAI steunt deze resolutie.

Dual-use

AI is bovendien een *dual-use* technologie, die zowel met goede als met kwade intenties kan worden ingezet. Voorkomen dient te worden dat AI in handen valt van mensen of regimes, die als doel hebben het in te zetten voor onderdrukkende, maatschappij-ontwrichtende of terroristische activiteiten.



SyRI

Met de Systeem Risico Indicatie-wet kreeg het Ministerie van Sociale Zaken de mogelijkheid om diverse soorten bestanden met gegevens van burgers te koppelen en doorzoeken voor het opsporen van fraude met belastingen of sociale uitkeringen. Door middel van risicomodellen wordt vastgesteld of er een verhoogd risico is op onregelmatigheden. Elk risicomodel bestaat uit vooraf bepaalde indicatoren, die per SyRI-systeem door de minister worden vastgesteld.

Al zoekt voorts verbanden in al die gegevens en doet een voorspelling. Het systeem doet dat echter niet op basis van de inhoudelijke merites van de kwestie. Het signaleert slechts verbanden (correlaties) met andere zaken waarin sprake is geweest van fraude.

Het kan de voorspellingen bovendien niet feitelijk, laat staan juridisch deugdelijk motiveren. Het bestuursorgaan kan dus ook zelf niet achterhalen waarom het systeem in een bepaald geval concludeert dat er sprake is van een risico op fraude en kan daarop gericht handelen dan ook niet motiveren.

De SyRI-wet is onlangs door de Rechtbank Den Haag in strijd met art. 8 EVRM beoordeeld. De Rechtbank vond dat er geen 'fair balance' bestaat tussen het doel dat met de wet wordt nagestreefd (voorkomen van fraude met overheidsregelingen) en het recht op een priveleven van art. 8 EVRM.

AI & Wet- en Regelgeving

AI opereert niet in een wetteloze wereld

Bij technologische vooruitgang komt altijd de vraag op of en zo ja, hoe we de nieuwe disruptieve technologie moeten reguleren. Zo ook bij AI. AI is echter geen volledig ongereguleerde technologie. We beschikken over vele wettelijke kaders die net zo goed op AI van toepassing zijn als op iedere andere technologie.

AI & Mensenrechten

AI is een technologie voor algemene doeleinden die invloed heeft op de hele samenleving, wat betekent dat de conceptie, ontwikkeling en toepassing van AI van invloed is op alle bestaande mensenrechten, op democratie en de rechtsstaat. Deze effecten kunnen zowel negatief als positief zijn, maar de impact van AI op deze rechten is tot nu toe negatiever dan positief.

De volgende 'families van mensenrechten' zijn in het bijzonder relevant voor de ontwikkeling en het gebruik van AI:

- Respect voor de menselijke waardigheid: in het kader van AI betekent dit dat alle mensen moeten worden behandeld als morele subjecten, en niet als ontleedbare, cijferbare, sorteerbare, conditioneerbare, manipuleerbare objecten.
- Vrijheid van het individu: in het kader van AI betekent dit dat mensen vrij zijn van (in)directe illegale dwang, bedreigingen voor de mentale autonomie en geestelijke gezondheid, ongerechtvaardigd toezicht, misleiding en oneerlijke manipulatie.
- Gelijkheid, non-discriminatie en solidariteit: AI-systeem mogen geen biased uitkomsten hebben.
- Sociale en economische rechten, zoals het recht op waardigheid op het werk en een veilige werkplek.

Democratie en Rechtstaat

Bovendien mogen democratische processen en de beginselen van de rechtsstaat niet worden ondermijnd door AI.

AI & Aansprakelijkheidsrecht

ALLAI is fel gekant tegen enige vorm van rechtspersoonlijkheid voor AI-systemen (net als het EESC en de AI HLEG). Van het aansprakelijkheidsrecht gaat een preventieve werking uit, die verdwijnt zodra de ontwikkelaar niet langer het aansprakelijkheidsrisico draagt.

Bestaande aansprakelijkheidsregels zijn bovendien ook van toepassing op AI. De EU Expert Group on Liability and AI heeft onlangs een rapport uitgebracht

AI en de GDPR/AVG

De GDPR (*General Data Protection Regulation*, in Nederland omgezet in de AVG) biedt enkele kaders voor het recht op uitleg van AI-besluiten (in de overwegingen) en het recht op menselijke tussenkomst.

Dit is echter onvoldoende om adequate transparantie in automatische besluitvorming te waarborgen, met name omdat ook een besluit dat op niet-persoonlijke gegevens is gebaseerd grote impact kan hebben.

AI & Consumentenrecht

Bestaande consumentenwetgeving is van toepassing, maar beschermt niet voldoende tegen oneerlijke, misleidende of manipulatieve praktijken (gepersonaliseerde prijsvorming, profilering, *nudging*).

Overig

Er bestaan er talloze sector-, domein- en gebruikersspecifieke wetten, regels en zelfregulering relevant voor AI. Denk aan het arbeidsrecht, bestuursrecht, IE-recht, vervoersrecht en gezondheidsrecht.

Facial Recognition

Gezichtsherkenning met AI is aan een snelle opmars bezig. Zo snel dat de technologie in diverse Amerikaanse steden is verboden of beperkt. De positieve applicaties van gezichtsherkenning, zoals het vinden van een vermist kind of een terrorist, zijn potentieel aanwezig, maar vooralsnog heeft de technologie een groot aantal negatieve effecten.

Het ongereguleerd en indiscriminatoir verzamelen van gezichten in een database kan het recht op privacy ondermijnen maar bovendien ook leiden tot een 'chilling effect', waarbij burgers zich constant bespied voelen en zich daar al dan niet bewust naar gaan gedragen. Ook kan het het recht op vereniging, staking of demonstratie ondermijnen, zo bleek tijdens de protesten in Hong Kong, waar demonstranten hun gezichten bedekken en lasers gebruiken om camera's met gezichtsherkenning te 'neutraliseren'.

Daar komt bij dat gezichtsherkenning verre van perfect is. Zo herkent het blanke mannen beter dan blanke vrouwen en is het nauwkeuriger in het herkennen van personen met een lichte huidskleur.

Ook wordt gezichtsherkenning (en overigens ook andere vormen van biometrische herkenning) meer en meer ingezet met de belofte dat AI de gemoedstoestand en emoties van mensen zou kunnen herkennen. Deze beloftes berusten echter op zeer wankel wetenschappelijk onderzoek en volgens sommigen zelfs op frenologie. Bovendien kan het het recht op mentale integriteit (onderdeel van het recht op een privacy) ondermijnen.

Wat is er nodig op het gebied van wet- en regelgeving?

Kennis, naleving en handhaving

AI-ontwikkelaars en (professionele) gebruikers van AI-systemen moeten begrijpen welke wet- en regelgeving van toepassing is op AI-systemen. Naleving van deze wet- en regelgeving moet voorop staan bij ontwikkelaars en professionele gebruikers van AI. En ten slotte, als regels niet worden gevolgd, moeten deze correct worden gehandhaafd. Dit is allemaal niet nieuw. Door de eeuwen heen hebben publieke en private instellingen te maken gehad met productvereisten, compliance-kwesties en handhaving van deze vereisten in verschillende vormen.

AI wet- en regelgeving stress-test

Het voorgaande geeft al een voorschot op de derde noodzakelijke stap: een AI stress-test voor bestaande wet- en regelgeving. Drie vragen moeten daarbij met name worden beantwoord:

- In hoeverre worden bestaande wettelijke doelstellingen beïnvloed door AI-systemen en op welke manieren?
- Zijn de bestaande kaders voor monitoring, informatievergaring en handhaving van de wetgevende maatregelen effectief?
- Werkt bestaande wetgeving zodanig dat zij de ethische principes en vereisten voor AI (zoals beschreven in de Ethische Richtlijnen voor Betrouwbare AI van de AI HLEG) bevorderen en waarborgen?

Risico- en voorzorgsbeginsel

Het karakter, de intensiteit en de timing van (aanpassing van) regelgeving moeten gerelateerd zijn aan het type en de mate van risico dat door een AI-systeem wordt gecreëerd.

Voor specifieke AI-toepassingen die "onaanvaardbare" risico's genereren of substantiële bedreigingen vormen, moet in plaats daarvan een op het voorzorgsbeginsel gebaseerde aanpak worden gevolgd.

Sluit aan bij EU stress-test activiteiten

ALLAI wijst erop dat de AI HLEG de Europese Commissie heeft geadviseerd om deze 'stress-test' op EU niveau uit te voeren, om zo Europa-breed tot een eenduidig regelgevend kader te komen. ALLAI adviseert de Nederlandse wetgever om aansluiting te zoeken bij een eventueel Europees stress-test proces, om zo dubbel werk en divergerende resultaten te voorkomen. Hetzelfde geldt voor (mogelijke) activiteiten van de Raad van Europa op het gebied van mensenrechten en AI.

Betrek alle ministeries

Om 'silo-denken' te voorkomen adviseert ALLAI eveneens om te zorgen voor samenwerking tussen alle relevante ministeries.

Generieke versus sector- of domeinspecifiek

In contexten die worden gekenmerkt door snelle technologische veranderingen, verdient het vaak de voorkeur om een principiële aanpak te hanteren, evenals een op resultaten gebaseerd beleid, mits met passend toezicht en handhaving.

Overweeg daarnaast sector- of domeinspecifieke wet- en regelgeving bij de verdere ontwikkeling van het regelgevingskader voor specifieke toepassingen van AI.

STRESS TEST

Nieuwe wet- en regelgeving voor (gebruik van) bepaalde AI-toepassingen

Een aantal AI-toepassingen is dermate intrusief of (potentieel) ontwrichtend dat nieuwe wetgeving op zijn plaats is. Onderstaande (niet limitatieve) aanbevelingen zijn gelijk aan de aanbevelingen van de AI HLEG. In de toekomst zullen er echter mogelijk nieuwe AI-toepassingen worden ontwikkeld waarvoor wetgeving moet worden overwogen.

Massa-surveillance

Onevenredige en massale surveillance van personen met gebruik van AI (bijvoorbeeld gezichtsherkenning), zowel door publieke als private partijen, moet worden voorkomen. Dergelijke systemen **maken** niet alleen een grote inbreuk op de privacy, maar hebben ook een 'chilling effect' op burgers en kunnen andere fundamentele rechten ondermijnen, zoals het recht op demonstratie en vereniging.

Biometrische herkenning

Personen mogen niet worden onderworpen aan ongerechtvaardigde persoonlijke, fysieke of mentale tracking of identificatie, profilering en nudging via AI-methoden voor biometrische herkenning zoals: emotionele tracking, empathische media, DNA, iris- en stemidentificatie, affectherkenning, gezichtsherkenning en de herkenning van micro-expressies.

Social scoring

Massale sociale scoring (zoals dat in delen van China wordt getest) moet worden voorkomen, maar ook beperkte scoring mechanismes (bijv. bij leningen) kunnen onze waarden en autonomie in gevaar brengen. Een volledig transparante procedure, voorafgaande toestemming (opt-in) of opt-out zijn daarbij onvoldoende waarborgen.

Verplichte zelfidentificatie van AI

Een AI-systeem mag zich niet heimelijk voordoen als mens.

Autonome wapens

Het Europees Parlement heeft opgeroepen tot de urgente ontwikkeling van een juridisch bindend standpunt ten aanzien van autonome wapens. De AI HLEG beveelt aan dat de ontwikkeling van geautomatiseerde dodelijke wapens moet worden bewaakt en beperkt.

Schone data-lei voor kinderen

De invoering van een wettelijke leeftijd waarop kinderen een 'schone datalei', geschoond van openbare of particuliere data die aan hen als kinderen zijn gerelateerd, krijgen, moet worden overwogen.



AI Research & Innovatie

AI kan niet langer worden gezien als een technische of informatica-discipline, het is per definitie multidisciplinair.

Multi-disciplinariteit

Gerichte maatregelen en steun voor onderzoek naar sociaal-maatschappelijke uitdagingen rondom AI en naar de ontwikkeling van veilige en robuuste AI-systemen zijn cruciaal. AI gaat over veel meer dan coderen en data alleen. De vele ethische, juridische en sociaal-maatschappelijke implicaties ervan vragen om een diverser en multidisciplinair onderzoek dan puur en alleen de mathematische kant van AI.

Mogelijke ongewenste sociaal-maatschappelijke implicaties van AI zouden kunnen worden geadresseerd door door gelijktijdig met financiering van de ontwikkeling van disruptieve AI-innovaties onderzoek te financieren naar de sociaal-maatschappelijke, ethische en juridische impact van juist die innovaties.

Een goed voorbeeld van een bestaande multidisciplinaire samenwerking is de Research Priority Area (RPA) Human(e) AI aan de Universiteit van Amsterdam, die zich richt op de maatschappelijke gevolgen van de snelle ontwikkeling van kunstmatige intelligentie (AI) en geautomatiseerde besluitvorming (ADM) in een breed scala van maatschappelijke gebieden. Human (e) AI is gebaseerd op samenwerking tussen de faculteiten wetenschap, rechten, geesteswetenschappen en sociale en gedragswetenschappen.

Diversiteit

Binnen de AI-wetenschappen moet bovendien diversiteit aangemoedigd worden. Meer vrouwen en mensen met verschillende achtergronden, inclusief mensen met een handicap, moeten worden betrokken bij de ontwikkeling van AI, om ervoor te zorgen dat AI niet-discriminerend en inclusief is.

SDG's

AI kan een uitgelezen mogelijkheid bieden om oplossingen te vinden voor urgente uitdagingen waarmee de samenleving te maken heeft, zoals een vergrijzende bevolking, toenemende sociale ongelijkheid en milieuvervuiling. Deze potentie is ook wereldwijd zichtbaar, bijvoorbeeld in de duurzameontwikkelingsdoelstellingen van de VN.

Investeer slim

Investerings in AI moeten slim zijn, gericht op domeinen en sectoren waar Nederland nu al sterk in is en echt een verschil kan maken, zoals 'Water en Maritiem', 'Agri en Food', 'Life Sciences & Health' en 'High-tech Systemen en Materialen'. In deze sectoren kan AI voor economische, sociaal-maatschappelijke en milieu voordelen zorgen.



"There's no AI-race, and if there is, it's the wrong one to run"

Vorgens Professor Virigina Dignum gokken met name de VS en China op ontwikkelingen op het gebied van machine learning en dan in het bijzonder deep learning: "Deze benaderingen richten zich echter op één aspect van intelligentie, dat van het vermogen om patronen waar te nemen en voorspellingen te doen op basis van die percepties. Echte intelligentie is meer dan dat en omvat de mogelijkheden om te redeneren, communiceren en beslissen op basis van kleine, onvolledige en tegenstrijdige informatie. Er is dringend behoefte aan alternatieven voor statistische benaderingen van leren. In feite heeft deze week, een studie die 25 jaar AI-onderzoek analyseert, geconcludeerd dat het tijdperk van diep leren ten einde komt.

Europa is traditioneel altijd sterk geweest op symbolische benaderingen van AI en op (sociale) robotica. Dit zijn precies enkele van de gebieden waarop momenteel moet worden geïnvesteerd, de technologieën die AI in de nabije toekomst vooruit zullen brengen. Het zou daarom een vergissing zijn om de VS en China blindelings te volgen op hun machine learning 'race' wanneer het nu de gelegenheid is om de waarde van alternatieve benaderingen aan te tonen. Benaderingen waarin wij Europeanen een voordeel kunnen hebben."

Tot slot

Tot slot wil ALLAI nog een aantal zaken meegeven die van belang zijn bij de beleidsbepaling rondom AI.

Het wiel hoeft niet opnieuw uitgevonden

Inmiddels is er bijzonder veel werk verzet en informatie beschikbaar over hoe de kansen van AI ten volle te benutten en de uitdagingen adequaat en tijdig het hoofd te bieden. Nederland hoeft het wiel dus niet opnieuw uit te vinden. Zij kan voortbouwen op het vele werk dat inmiddels met name op EU-niveau is verzet en gebruikmaken van de expertise die bijvoorbeeld bij ALLAI beschikbaar is.

Het is met name niet nodig om (publiek) geld en tijd te steken in het opstellen van (weer) een set ethische richtlijnen. Het is nu zaak om een inhaalslag te maken, in de eerste plaats op het gebied van kennis en begrip van de kansen en uitdagingen van AI zoals omschreven in dit Position Paper. Daarnaast moet Nederland werk maken van de implementatie van de Ethische Richtlijnen van de AI HLEG, met name op publiek niveau. Ten slotte moet Nederland zich snel op de hoogte stellen van en aansluiting zoeken bij de huidige ontwikkelingen op EU en internationaal niveau op het gebied van AI-beleid en wetgeving.

Geen AI-race

Pers en beleidsmakers waarschuwen vaak voor een zogenaamde AI-race en het feit dat Nederland en Europa die race aan het verliezen zouden zijn. Naar de mening van ALLAI is dit discours zowel verkeerd als gevaarlijk.

Vaak vergeten dat Europa de grootste markt van de wereld is met 500 miljoen consumenten. Dit geeft de EU de mogelijkheid om de randvoorwaarden te stellen aan AI en daarmee ook wereldwijd invloed te hebben.

Een goed voorbeeld van een dergelijke invloed is de GDPR. Deze verordening maakt dat bedrijven van buiten de EU zich aan strenge data privacy eisen moeten houden, maar vindt inmiddels ook gehoor bij beleidsmakers in die andere delen van de wereld.

Betrek de brede maatschappij

Vanuit de Nederlandse traditie wordt bij beleidsvorming en uitvoering vaak gekozen voor een publiek-private benadering op basis van de "triple helix" (overheid, bedrijven, kennisinstellingen). ALLAI is van mening dat de impact van AI zo breed is dat er gekozen zou moeten worden voor een "full-helix-benadering", waarbij de maatschappij in de breedste zin is betrokken.

Dit betekent bedrijven, onderzoekers uit diverse disciplines (waaronder AI-wetenschap, data-science, ethiek, filosofie, recht, economie en met name ook sociale wetenschappen (zoals culturele antropologie, pedagogiek, politicologie en psychologie), centrale en decentrale overheden en maatschappelijke organisaties, inclusief de sociale partners.

Bronnen

- VPRO Tegenlicht, Mens in de Machine
- EESC INT/806, Artificial Intelligence & Society, 2018, Rapporteur: Catelijne Muller
- Europese Commissie, Artificial Intelligence for Europe, COM(2018) 237 final
- Europese Commissie, Coordinated Plan on Artificial Intelligence (COM(2018) 795 final)
- Ethics Guidelines for Trustworthy AI, High Level Expert Group on AI, 2019
- Policy and Investment Recommendations for Trustworthy AI, High Level Expert Group on AI, 2019
- Von der Leyen: "A Union that strives for more: My agenda for Europe"
- Amazon ditched AI recruiting tool that favored men for technical jobs", The Guardian via Reuters, 11 October 2018
- Europese Commissie Communicatie, Building Trust in Human-Centric Artificial Intelligence (COM(2019) 168 final)
- De Robot de Baas, WRR, Robert Went et al.
- Netflix documentaire The Great Hack
- Resolutie 2018/2752(RSP) van het Europees Parlement.
- C. McCrudden, Human Dignity and Judicial Interpretation of Human Rights, *EJIL*, 19(4), 2008
- E. Hilgendorf, Problem Areas in the Dignity Debate and the Ensemble Theory of Human Dignity, in: D. Grimm, A. Kemmerer
- C. Möllers, *Human Dignity in Context. Explorations of a Contested Concept*, 2018, blz. 325 ff.
- Liability for Artificial Intelligence, 2019, EU Expert Group on Liability and New Technologies to the European Commission
- AINow 2019 Report
- ALLAI | AIREA-NL: Recommendations for the Artificial Intelligence Research Agenda for the Netherlands





ALLAI.

ALLAI verwijst naar Stichting ALLAI Nederland, een stichting naar Nederlands recht. Geen enkele entiteit of persoon gelieerd aan ALLAI is verantwoordelijk of aansprakelijk voor enig direct of indirect verlies of enige schade geleden door een entiteit of persoon die op deze communicatie vertrouwt.

© 2020 ALLAI, Nederland