Oude Waalsdorperweg 63
2597 AK Den Haag
Postbus 96864
2509 JG Den Haag

www.tno.nl

T +31 88 866 10 00
F +31 70 328 09 61

Aan
Dr. M.T. Croes – Ministerie van Justitie en Veiligheid
Dr. M.J.P. van Veen – Ministerie van Defensie
M. Emmelkamp – Ministerie van Buitenlandse Zaken

Van
Drs. R.M. Neef
C.E.A. van Weerd, MA

Onderwerp
Artificial Intelligence in de context van de nationale veiligheid – Eindnotitie
Studie binnen het Analistennetwerk Nationale Veiligheid

Datum
27 januari 2020

Onze referentie
TNO 2020 M10094

E-mail
carolina.vanweerd@tno.nl

Projectteam ANV/TNO
Drs. R.M. Neef
C.E.A. van Weerd, MA
H. Bonte, MA
I.N. Melman, BA
Ir. L. Gooijer
Dr. H.L. Duijnhoven

Inhoud

Datum
27 januari 2020

Onze referentie
TNO 2020 M10094

1 Introductie op de verkenning van AI in het kader van de nationale veiligheid

1.1 Inleiding

Het Analistennetwerk Nationale Veiligheid[1], is door de ministeries van Justitie & Veiligheid, Defensie en Buitenlandse Zaken gevraagd om een studie te doen naar de mogelijke risico's en dreigingen van Kunstmatige Intelligentie, (*Artificial Intelligence*, AI) toepassingen op de nationale veiligheid, waarbij interne en externe veiligheid met elkaar zijn verweven.[2] TNO, als één van de kernpartners binnen het netwerk, heeft deze studie uitgevoerd. Dit onderzoek is onderdeel van het vormgeven van een gezamenlijke onderzoeksagenda van de drie departementen rond de zogenoemde 'nexus interne en externe veiligheid'.

1.2 Aanleiding

Kunstmatige Intelligentie is het vermogen van systemen om intelligent gedrag te vertonen en omvat een brede familie van technologieën, geïnspireerd op het menselijk denken en handelen. AI vanaf de jaren '50 van de vorige eeuw worden algoritmes ontwikkeld die proberen om menselijke vaardigheden als redeneervermogen, beeld- en spraakherkenning en leervermogen te emuleren in computersystemen. Dit heeft geleid tot tal van toepassingen die we dagelijks om ons heen zien, zoals zoekmachines op het Internet, de chatbots van online winkels, filevoorspellers, drones en robots.

Figuur 1: AI in de headlines van Nederlandse media

[1] Het Analistennetwerk Nationale Veiligheid (ANV) is een kennisnetwerk dat sinds 2011 analyses maakt in het kader van de Nationale Veiligheidsstrategie. *Analistennetwerk Nationale Veiligheid* (website), https://www.rivm.nl/rivm/organisatie/centrum-veiligheid/analistennetwerk-nationale-veiligheid.

[2] Deze verwevenheid van interne en externe veiligheid is in meerdere (beleids)documenten opgemerkt, waarbij de kernpremisse zich centreert rondom de gedachte dat "[g]ebeurtenissen buiten Nederland [immers] significante gevolgen [kunnen] hebben voor de (organisatie) van veiligheid in Nederland." NCTV, *Nationale Veiligheid Strategie 2019*, 9.

Ontwikkelingen in het AI vakgebied zorgen wereldwijd voor veel aandacht, in Figuur 1 een kleine greep uit de *headlines* van Nederlandse media.[3] AI-technologie biedt veel kansen, onder andere in het economische domein, maar toepassing van AI-technologie in specifieke systemen brengt ook potentiële risico's en dreigingen met zich mee. Dergelijke risico's en dreigingen zijn deels nog onbekend en moeilijk te voorspellen, waardoor met enige regelmaat verhalen circuleren op het internet die op dit moment onwaarschijnlijk lijken, zoals het idee dat robots de wereld overnemen.[4]

Ondanks dat we nog niet goed de potentiële risico's en dreigingen van AI-technologie kunnen inschatten, zal deze zeer waarschijnlijk de nationale veiligheid gaan raken. Om te begrijpen op welke manieren de nationale veiligheid potentieel geraakt kan worden is het noodzakelijk om beter zicht te krijgen op het begrip AI en de context waarin AI-technologie wordt toegepast. Binnen het vakgebied AI bestaat geen eenduidige definitie van de technologie. AI is een verzamelterm voor verschillende algoritmes, modellen en perspectieven. De producten die in de volksmond 'AI-systemen' worden genoemd kunnen wel kenmerkende eigenschappen bezitten, zoals zelflerend of herkennend vermogen, maar kunnen allerlei vormen aannemen, en werken op basis van verschillende architecturen en algoritmes. Elk van die variaties kan specifieke risico's met zich meebrengen als deze toegepast worden in een product.

AI is dus een complex fenomeen en er kan niet direct worden gesproken over dé impact van dé technologie AI op dé nationale veiligheid. Om die reden is het nodig om eerst een gemeenschappelijk kader te creëren. Waar hebben we het over als het gaat om AI in de context van nationale veiligheid? Hierbij zijn niet alleen de

[3] Bauke Schievink, "AI van google kan 26 huidaandoeningen net zo goed herkennen als dermatologen," *Tweakers* 14 september 2019, https://tweakers.net/nieuws/157340/ai-van-google-kan-26-huidaandoeningen-net-zo-goed-herkennen-als-dermatologen.html; Laurens Verhagen, "'Regering moet veel meer investeren in kunstmatige intelligentie'," *de Volkskrant* 28 september 2018, https://www.volkskrant.nl/columns-opinie/regering-moet-veel-meer-investeren-in-kunstmatige-intelligentie~b2779cce/; Bennie Mols, "Slimme AI kan nóóit een mens worden," *NRC* 8 september 2019, https://www.nrc.nl/nieuws/2019/09/08/slimme-ai-kan-nooit-een-mens-worden-a3972635; Laurens Verhagen, "Ook cybercrimineel ontdekt aantrekkelijke kant van kunstmatige intelligentie – hoe zorgwekkend is dat?," *de Volkskrant* 11 januari 2019, https://www.volkskrant.nl/nieuws-achtergrond/ook-cybercrimineel-ontdekt-aantrekkelijke-kant-van-kunstmatige-intelligentie-hoe-zorgwekkend-is-dat~bb9327f8/; Johannes Fahrenfort, "Het summum van gevaar: onvoorspelbare kunstmatige intelligentie," *de Volkskrant* 20 november 2017, https://www.volkskrant.nl/columns-opinie/het-summum-van-gevaar-onvoorspelbare-kunstmatige-intelligentie~bbff588b/; "Microsoft en Amazon dreigen betrokken te raken bij productie killer robots," *MO Mondiaal Nieuws* 19 augustus 2019, https://www.mo.be/nieuws/microsoft-en-amazon-dreigen-betrokken-te-raken-bij-productie-killer-robots; "Microsoft steekt een miljard in AI-onderzoek: 'Kan de koers van de mensheid bepalen'," *AD* 23 juli 2019, https://www.ad.nl/tech/microsoft-steekt-een-miljard-in-ai-onderzoek-kan-de-koers-van-de-mensheid-bepalen-br~afc78535/?referrer=https://www.google.com/; Thomas Riemens, "'Mens kan controle over killer robots binnen paar jaar kwijt zijn'," *NOS* 15 november 2017, https://nos.nl/artikel/2202758-mens-kan-controle-over-killer-robots-binnen-paar-jaar-kwijt-zijn.html; "AI kan leiden tot 12-urige werkweek, denkt Alibaba-miljardair Jack Ma," *Nu.nl* 30 augustus 2019, https://www.nu.nl/economie/5985631/ai-kan-leiden-tot-12-urige-werkweek-denkt-alibaba-miljardair-jack-ma.html.

[4] Dave Partner, "Blockchains: The AI that will take over the world has already been born!" *Medium* 17 november 2017, https://becominghuman.ai/blockchains-the-ai-that-will-take-over-the-world-has-already-been-born-73041e759c8.

technische aspecten van belang, maar ook het totale ecosysteem waarin deze technologie bestaat.

1.3 Probleem- en doelstelling

De centrale probleemstelling van dit onderzoek is:

Wat is Kunstmatige Intelligentie (Artificial Intelligence, AI), welke vormen, fenomenen, actoren en factoren spelen een rol bij de ontwikkeling en adoptie daarvan, en hoe verhoudt zich dat tot de nationale veiligheid?

Bovenstaande probleemstelling is in twee delen op te delen: aan de ene kant het duiden van het begrip AI (wat is het 'speelveld' van dit vakgebied?) en aan de andere kant het duiden van de verhouding van deze technologie tot de nationale veiligheid.

Het eerste deel van de probleemstelling richt zich op het duiden van AI. Hoewel er, zoals aangegeven, geen algemeen geaccepteerde definitie van AI bestaat wordt hier uitgegaan van de algemene omschrijving van AI als brede familie van technologieën geïnspireerd op het menselijk denken en handelen, en gericht op het ontwikkelen van het vermogen van systemen om intelligent gedrag te vertonen. Om de probleemstelling te kunnen beantwoorden, moet vanuit deze algemene omschrijving gekeken worden naar fenomenen die zijn verbonden met dit technologiegebied. Hierbij moet niet alleen gekeken worden naar de verschillende technieken en algoritmes die onder de verzamelterm AI worden gevat, maar ook naar de manier waarop dergelijke technieken of modellen worden ingezet om bepaalde functies te vervullen (zoals patroonherkenning in grote hoeveelheden informatie) ten behoeve van bepaalde toepassingen (producten). Deze producten kennen weer specifieke technische randvoorwaarden (zoals voldoende beschikbare data) en worden geconfronteerd met bepaalde maatschappelijke contexten (denk aan publiek vertrouwen, maar ook wetgeving). Ook de economische context speelt een rol, evenals de actoren die een AI-systeem gebruiken, dan wel ontwikkelen. Door al deze verschillende fenomenen met betrekking tot AI te structureren en de onderlinge verbinding te expliciteren (in dit geval in de context van nationale veiligheid), ontstaat een (tijdelijk) gezamenlijk denkkader. Tijdelijk, omdat, afhankelijk van de focus die je neemt, de inhoudelijke verzameling van aan AI verbonden concepten, en de nadruk die ze krijgen, anders kunnen zijn. Ook is het onderwerp continu in beweging dus zal een gestructureerd kader door de tijd veranderen.

Het tweede deel van de probleemstelling draait om het plaatsen van het speelveld van AI in de context van de nationale veiligheid. Hierbij wordt gekeken naar mogelijke risico's en dreigingen[5] voor de nationale veiligheid die voortvloeien uit de toepassing van AI-technieken. Conform de Nationale Veiligheidsstrategie (NVS) is de nationale veiligheid in het geding als één of meer van de zes nationale

[5] Hoewel toepassing van AI zowel kansen als risico's kent, wordt in de context van de nationale veiligheid vooral gekeken naar de mogelijke negatieve impact op de samenleving die kan leiden tot aantasting van één of meer van de nationale veiligheidsbelangen zoals die zijn gedefinieerd in de Nationale Veiligheidsstrategie (NVS). Het gaat om (i) territoriale veiligheid, (ii) fysieke veiligheid, (iii) economische veiligheid, (iv) ecologische veiligheid, (v) sociale en politieke stabiliteit en (vi) internationale rechtsorde.

veiligheidsbelangen zodanig bedreigd worden dat sprake is van (potentiële) maatschappelijke ontwrichting. Net als AI is nationale veiligheid een dynamisch en veelzijdig begrip, dat op verschillende wijzen kan worden aangetast.

Om de twee delen van de vraagstelling (AI en nationale veiligheid) met elkaar te kunnen verbinden is het belangrijk om bij de beantwoording van de vraag wat AI is ook in te gaan op mogelijke afhankelijkheden, kwetsbaarheden en misbruik van AI-toepassingen. Uit het bovenstaande vloeit de volgende doelstelling van het project:

Het in kaart brengen van het fenomeen AI ('de zin van de onzin scheiden'), afhankelijkheden, kwetsbaarheden en mogelijk misbruik die gepaard gaan met de brede introductie van AI en de potentiele risico's en dreigingen daarvan in de context van de nationale veiligheid.

1.4 Leeswijzer

In hoofdstuk 2 van deze notitie wordt de aanpak die is gehanteerd in dit project toegelicht. In hoofdstuk 3 en 4 wordt vervolgens ingegaan op de inhoudelijke uitkomsten van deze verkenning, aan de hand van een '*concept map'* (een kennisrepresentatie over het thema AI). De bruikbaarheid van deze concept map wordt eveneens in hoofdstuk 3 toegelicht. In hoofdstuk 4 wordt ingegaan op mogelijke risico's en dreigingen van AI-toepassingen op de nationale veiligheid, waarna in hoofdstuk 5 wordt ingegaan op mogelijke vervolgonderzoeksvragen volgend uit deze verkenning. In hoofdstuk 6 wordt ingegaan op de bruikbaarheid van de in dit project ontwikkelde aanpak als (generieke) methodiek om andere nieuwe technologieën en technologische ontwikkelingen te duiden. Hoofdstuk 7 geeft tot slot een conclusie. In de bijlagen staan een overzicht van relevant bronmateriaal (Bijlage 1), een lijst met organisaties van de voor dit project geraadpleegde experts (Bijlage 2) en een weergave van de volledige concept map verkenning Kunstmatige Intelligentie in de context van nationale veiligheid (Bijlage 3).

2 Aanpak

In dit project zijn verschillende stappen doorlopen om AI in de context van nationale veiligheid te duiden. Als eerste is aan de hand van een kennisrepresentatie over AI (in de vorm van een *concept map*) gewerkt aan een bruikbare duiding van het begrip AI. Vervolgens is een eerste inzicht ontwikkeld in mogelijke dreigingen en risico's die gepaard kunnen gaan met de toepassing van AI-technologie. De gevolgde aanpak in dit project zal hieronder uiteengezet worden. In hoofdstuk 6 wordt op basis van de gevolgde aanpak in het project een generieke methodiek toegelicht om nieuwe technologieën en technologische ontwikkeling te kunnen duiden.

Zoals in de inleiding is aangegeven, richt het eerste deel van de centrale doelstelling van deze verkenning zich op het in kaart brengen van AI als technologiegebied, met alle facetten die daarbij horen. Ten behoeve van een eerste ordening van AI is gestart met een literatuurverkenning. De bronnen zijn geselecteerd op basis van een aantal criteria:

- Openbaar beschikbaar;
- Recentelijk uitgebracht (concreet: niet ouder dan 2017);
- Connectie tussen AI en (nationale) veiligheid (voor zover af te leiden van de titel).

Om enige diversiteit aan bronmateriaal te creëren, hebben we aanvullend op bovenstaande criteria gezocht op verschillende 'typen' bronnen: 1) overheid, 2) bedrijfsleven en innovatie, 3) denktanks en analyse en 4) wetenschap. Daarnaast hebben we gekeken naar bronnen binnen Nederland, binnen Europa en andere internationale bronnen.[6]

Het doel van het project was niet om een uitputtende literatuurstudie uit te voeren. Voor het maken van een eerste (gedegen) versie van de ordening van het technologiegebied AI was een eerste verkenning van bestaande literatuur echter wel relevant. Het projectteam heeft in meerdere werksessies vanuit de geraadpleegde bronnen een basisopzet gemaakt voor een ordening. Het doel hiervan was om relevante concepten op het gebied van AI op een hoger abstractieniveau te clusteren en indien mogelijk met elkaar te verbinden. Op deze manier is een eerste indeling gemaakt in hoofdconcepten: randvoorwaarden, sectoren, toepassingen, methoden en technieken en actoren. Deze thema's kwamen in vrijwel elke bron terug, wat indiceert dat het thema een relevante plek heeft in het vakgebied van AI-technologie.

De keuze voor de specifieke terminologie van de thema's (hoofdconcepten) is tot stand gekomen door de concepten die veel in de literatuur naar voren kwamen op een hoger aggregatieniveau te clusteren. Wanneer bijvoorbeeld in verschillende artikelen werd gesproken over China, de Verenigde Staten, Google, Huawei en universiteiten als belangrijke spelers op het gebied van AI-technologieontwikkeling hebben wij als projectteam ervoor gekozen deze concepten te clusteren onder het hoofdconcept 'actoren'. Bedrijven als Google en Huawei zijn binnen het hoofdconcept actoren geclusterd in het concept 'technologiebedrijven'.

[6] In bijlage 1 staat een overzicht van relevant bronmaterieel ingedeeld in de vier 'typen' bronnen.

Technologiebedrijven zijn vervolgens naar aanleiding van een expertsessie uitgesplitst op thema (wat doet een dergelijk bedrijf?). Dit geeft meer inzicht in het speelveld van AI dan een opsomming van de op dit moment relevante technologiebedrijven. Uit deze clustering vanuit de geraadpleegde literatuur is een eerste indeling van hoofdconcepten opgesteld door het projectteam, waarna dit in de volgende stappen met de experts is besproken, verder uitgewerkt en vastgesteld.

Voor het ordenen, vastleggen en verrijken van de kennisrepresentatie rondom AI is binnen dit project de methode *concept mapping* gebruikt. In deze methode worden 'concepten' visueel op een kaart gezet en middels verbindingen hun relaties aangeduid. Concept mapping als methode heeft geholpen om concepten te clusteren en verbanden tussen de (hoofd)concepten te expliciteren. In dit project is, zoals gezegd, deze kaart iteratief opgebouwd door het projectteam en experts tijdens interactieve werksessies. Door middel van interne werksessies is een eerste versie van de concept map gecreëerd die is aangevuld en bijgewerkt in een expertsessie met technisch-inhoudelijke experts op het gebied van AI-technologie.[7] In de sessie zijn 'live' de wijzigingen doorgevoerd in de concept map en na de sessie verder uitgewerkt. Ook in de tweede expertsessie, met beleidsmedewerkers van de ministeries van Justitie & Veiligheid, Defensie, Buitenlandse Zaken en Binnenlandse Zaken en Koninkrijksrelaties, is de concept map toegelicht en aangevuld. In hoofdstuk 3 wordt het gebruik van de concept map verder toegelicht.

Het tweede deel van de doelstelling van dit project draaide om het in kaart brengen van de potentiële risico's en dreigingen die gepaard kunnen gaan met de toepassing van AI-technologie. Vanuit de structurering van het 'speelveld' AI is vervolgens een eerste stap gemaakt naar mogelijke risico's en dreigingen. De AI-concept map diende hierbij als uitgangspunt. Verschillende combinaties van concepten op de concept map leveren eerste indicaties van mogelijke risico's en dreigingen, inclusief aanknopingspunten voor de context waarbinnen een risico geduid kan worden. Concreet kunnen bijvoorbeeld combinaties gemaakt worden tussen een actor (een staat, bijvoorbeeld China), een toepassing van AI (onbemande autonome wapensystemen) en een sector (de veiligheidssector). Deze combinatie van concepten leidt dan tot een 'dreigingsnarratief': China gebruikt (in de toekomst) onbemande autonome wapensystemen op het gevechtsveld. Een dergelijk specifieker narratief biedt concrete aanknopingspunten voor het bepalen van de potentiële impact op nationale veiligheid.

In dit project is ervoor gekozen om eerst algemene risicocategorieën te definiëren (archetypische risico's rondom de technologie in kwestie, zoals 'een AI-gestuurd systeem neemt een verkeerde beslissing'), waarna er binnen deze categorieën meer concrete instanties van deze risico's en dreigingen geïnventariseerd kunnen worden (bijvoorbeeld 'een autonome auto veroorzaakt ongeluk door verkeerde beeldinterpretatie'). In beide gehouden expertsessies is, naast de concept map, ook aandacht besteed aan de mogelijke risico's en dreigingen van AI-toepassingen.

[7] Zie bijlage 2 voor een lijst met organisaties van geraadpleegde experts binnen dit project.

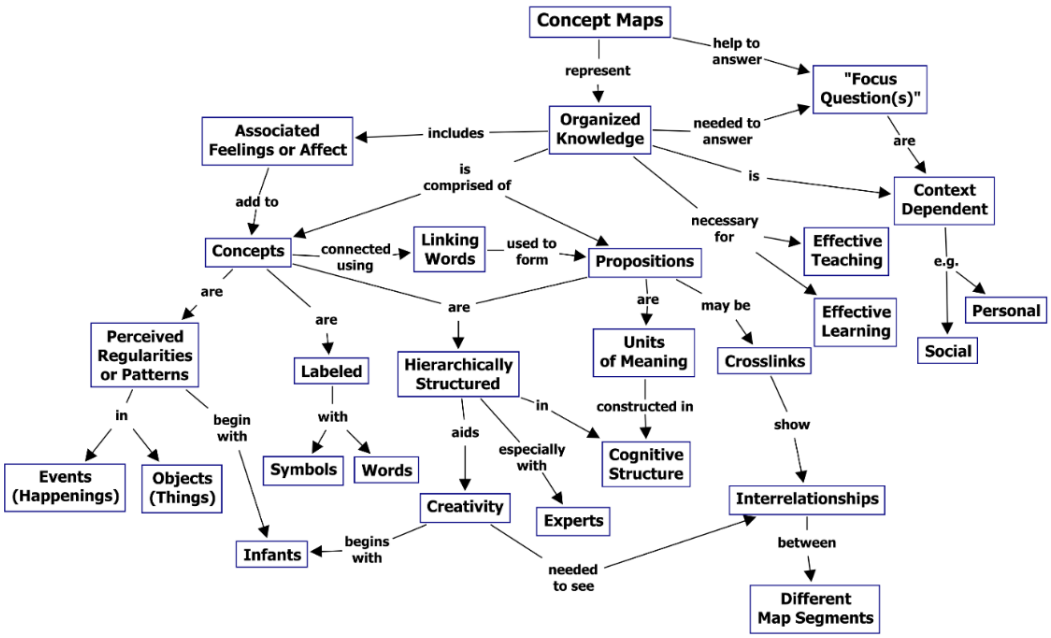
In de volgende hoofdstukken wordt ingegaan op de resultaten van de verkenning: de concept map waarmee AI in de context van nationale veiligheid in kaart is gebracht en eerste indicaties van mogelijke risico's en dreigingen die gepaard kunnen gaan met de toepassing van AI-technologie.

3 Het in kaart brengen van AI middels concept mapping

3.1 Concept mapping als methode

Een cruciaal onderdeel van dit project is het identificeren van elementen die in verband gebracht worden met AI. Hiervoor is een methode gebruikt die deze elementen letterlijk in kaart brengt: concept mapping. Concept mapping komt voort uit de cognitieve psychologie en is ontwikkeld om kennisstructuren weer te geven zoals mensen deze opbouwen door te leren.[8]

Een concept map is een visuele weergave van onderwerpen, waarbij de verschillende concepten binnen het onderwerp zich op een bepaalde manier tot elkaar verhouden (zie Figuur 2[9]). Deze onderlinge verbanden worden in een concept map expliciet benoemd door middel van een woord (of een aantal woorden) op de pijl die de concepten met elkaar verbindt (of verbinden). Deze relaties zijn verschillend, concepten kunnen bijvoorbeeld gelijksoortig zijn of hebben een oorzaak/gevolg relatie. Concreet worden deze verbanden weergegeven als werkwoorden, zoals 'hangt samen met', 'ondersteunt', 'is onderdeel van', 'is een voorwaarde voor' en 'veroorzaakt'.[10]



Figuur 2: Voorbeeld van een conceptmap, waarin het fenomeen Concept Maps (bovenaan de map) wordt uitgelegd.

[8] Joseph D. Novak, Dismas Musonda, "A Twelve-Year Longitudinal Study of Science Concept Learning," *American Educational Research Journal*, 28, no. 1 (January 1991), 117-153.

[9] Alberto J. Cañas, Joseph D. Novak, "What is a Concept Map?," *CMap (website)*, http://cmap.ihmc.us/docs/conceptmap.php.

[10] https://www.kuleuven.be/onderwijs/werken_opo/conceptmapping_OPO.

Concept mapping is een manier om een conceptueel raamwerk te creëren voor plannings-, ontwikkelings- en evaluatiedoeleinden. Eerst moeten de basale concepten worden gegenereerd. Na de generatie van ideeën wordt een rangorde aangebracht en eventueel geclusterd.[11] De expliciete verbanden en een definitieve verzameling concepten worden bepaald in verschillende expertsessies. Het idee is dus dat de concept map op deze manier iteratief wordt opgebouwd.

Een concept map lijkt op een mind map, maar het verschil zit in het expliciteren van verbanden tussen concepten in een concept map, waar een mind map eerder een opsomming is van verschillende ideeën die een connectie hebben met het kernidee.

Het doel van het gebruik van de concept map in dit project is het (relatief snel) inzichtelijk kunnen maken van een complex begrip als AI in de context van nationale veiligheid, waarbij het niet de bedoeling is om een complete definitie van het fenomeen AI te maken. De concept map moet inzichten geven en aanknopingspunten bieden voor beleidsontwikkeling en onderzoeksprogramma's. De concept map biedt in deze zin een aanzet tot een inhoudelijke discussie en dwingt mensen om een brede blik te hanteren en niet (onbewust) al te verkokerd te kijken naar het concept (in dit geval AI). Bovendien helpt de concept map om op een andere manier na te denken over mogelijke risico's en dreigingen. De combinaties van verschillende concepten schetsen namelijk de context waarbinnen specifieke risico's en dreigingen kunnen ontstaan, waarbij de map opnieuw een bredere blik kan faciliteren. Een verdere toelichting op hoe de concept map AI gebruikt kan worden staat in paragraaf 3.3.

Concreet is in dit project gebruik gemaakt van een online beschikbare software tool ("*CmapTools*") van het *Institute for Human-Machine Cognition* (IHMC) in Florida.[12]

3.2 Duiding van de AI-concept map

De AI-concept map die is opgebouwd in dit project is gebaseerd op interne concept mapping exercities (waarbij gebruik is gemaakt van literatuur) en twee expertsessies. In deze activiteiten zijn concepten opgehaald, verbanden tussen concepten gelegd en ideeën uitgewisseld over de belangrijkste concepten in het technologieveld AI.

Een concept map is een levend document. Het is een gezamenlijk opgebouwde representatie van hoe er in de geraadpleegde literatuur en door de geraadpleegde experts naar AI gekeken wordt en daarmee biedt het een (tijdelijk) gevalideerd denkkader met betrekking tot AI. Het is van belang te benadrukken dat de concept map open staat voor additionele perspectieven. De concept map die in dit project is ontwikkeld, is te vinden in bijlage 3.

De concept map (Cmap) '*verkenning Kunstmatige Intelligentie in de context van nationale veiligheid'* is vormgegeven rondom een aantal hoofdcategorieën, weergegeven met de gekleurde cirkels in de Cmap. Deze concepten zijn tijdens de verkenning door clustering van materiaal geïdentificeerd als de belangrijkste

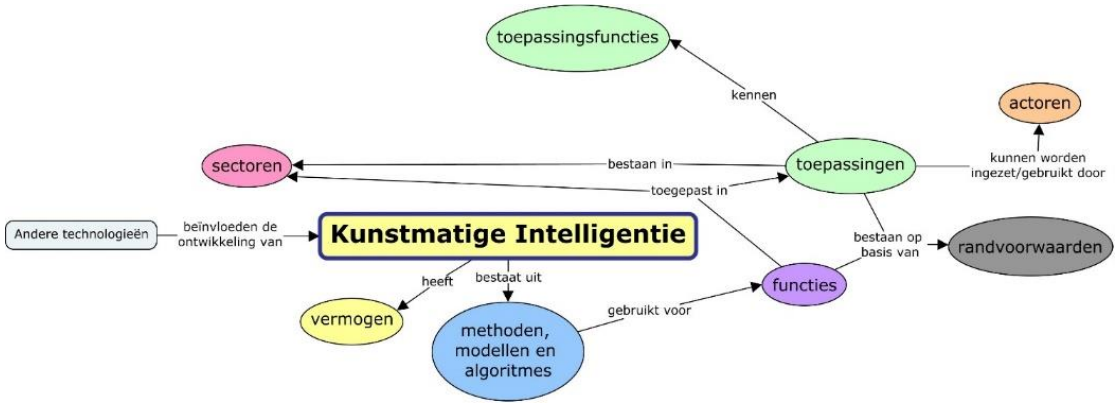
[11] Stephanie Sutherland en Steven Katz, "Concept mapping methodology: A catalyst for organizational learning," *Evaluation and Program Planning* 28, no. 3 (augustus 2005): 258, geraadpleegd op 2 juli 2019, https://doi.org/10.1016/j.evalprogplan.2005.04.017

[12] "Cmap" (*website*) https://cmap.ihmc.us/

dimensies binnen het technologieveld AI. De expertsessies zijn gebruikt om de hoofdindeling te valideren met experts en waar nodig aan te scherpen. Het mag duidelijk zijn dat de keuze voor deze concepten nauw verbonden is met de vraagstelling van het project (AI in de context van nationale veiligheid). Een andere projectvraag zou kunnen leiden tot alternatieve hoofdconcepten. In dit project zijn we gekomen tot de volgende zeven hoofdconcepten:

- **Vermogen:** Wat is de achterliggende doelstelling van de toepassing van AI en in welke mate emuleert het systeem menselijke intelligentie?
- **Methoden, modellen en algoritmes:** Welke verschillende technieken vallen er te onderscheiden binnen AI-technologie en hoe zijn die te typeren?
- **Functies:** Wat zijn generieke systeemfuncties op basis waarvan een AI-systeem functioneert?
- **Toepassingen en toepassingsfuncties:** Wat zijn karakteristieke toepassingen van AI-technologie, waar wordt een systeem voor ingezet?
- **Sectoren:** Wat zijn relevante sectoren waarin AI-technologieën toegepast worden?
- **Actoren:** Welke actoren gebruiken AI-toepassingen of zijn relevant in het vakgebied?
- **Randvoorwaarden:** Wat zijn de randvoorwaardelijke zaken die de toepassing van AI-technologie mogelijk maken, en beïnvloeden?

Figuur 3 toont de hoofdconcepten van de Cmap en hun relatie tot het centrale thema Kunstmatige Intelligentie. Omdat AI-technologie niet een geïsoleerde technologie is, is in de Cmap opgenomen dat de ontwikkeling van andere technologieën ook invloed kan hebben op AI als technologisch vakgebied.



Figuur 3: Cmap Kunstmatige Intelligentie - Hoofdconcepten

In de volgende paragrafen bespreken we de hoofdconcepten uit de Cmap, en worden belangrijkste onderliggende concepten uitgelicht (concepten op de map die worden besproken zijn onderstreept in de tekst).

3.2.1 Hoofdconcept: Vermogen

'Vermogen' als hoofdconcept (zie Figuur 5) gaat over de mate waarin een AI-systeem menselijke intelligentie emuleert. Van oudsher wordt er gesproken over verschillende klassen van systemen. Met de term '*Narrow* AI' (*Artificial Narrow Intelligence*, vaak afgekort als 'ANI') worden typisch systemen aangeduid die binnen afgebakende grenzen intelligent gedrag vertonen[13]. Dit zijn bijvoorbeeld systemen die zeer veel kennis hebben over een bepaald onderwerp, of een bepaalde cognitieve taak kunnen uitvoeren, zoals leren of patroon herkennen. Alle huidige AI-toepassingen vallen binnen deze categorie, van eenvoudige patroonherkenners, autonome platformen tot complexe lerende systemen. Ze vertonen intelligent gedrag, maar binnen afgebakende grenzen. Zelfs systemen die op het oog zeer intelligent, bijna natuurlijk gedrag vertonen, zijn nog steeds voorbeelden van 'narrow AI'. Denk bijvoorbeeld aan communicerende systemen zoals Google Assistant of Siri[14], de bewegende robots van Boston Dynamics[15] en verschillende humanoïde robots die er als mensen uitzien en menselijk praten. Deze, en alle andere AI-toepassingen vallen onder de typering 'Narrow AI' omdat ze geen daadwerkelijke wereldkennis hebben, geen diep begrip opbouwen en alleen kunnen werken binnen de kaders van het ontwerp. Om deze redenen, wordt 'Narrow AI' ook wel '*Weak* AI' of '*Soft* AI' genoemd: systemen die maar op een bepaald vlak in een bepaald terrein intelligent genoemd mogen worden.

Als systemen hogere cognitieve functies kunnen uitvoeren, en intelligentie vertonen op menselijk niveau, dan spreken we over '*Artificial General Intelligence*' (AGI), of '*Strong* AI'. Een AGI systeem heeft de potentie om elke willekeurige cognitieve taak uit te kunnen voeren, en dat kenmerkt zich in wereldbegrip, creativiteit en probleemoplossend vermogen. Er zijn op dit moment nog geen systemen die met goede onderbouwing een toonbeeld van AGI genoemd kunnen worden. Onze ideeën over AGI systemen worden vooral gevoed door toekomstbeelden uit *science fiction* waarin robots, computersystemen en mensen naadloos samenwerken, of waarin systemen proberen de wereld over te nemen. AGI kan gezien worden als *next step* na Narrow AI, maar de voorspellingen wanneer we volwaardige AGI applicaties gaan zien, lopen in de tientallen jaren. Daarnaast is er veel scepsis onder sommige wetenschappers en analisten of systemen überhaupt menselijke cognitie kunnen behalen. Dit is een langlopend cognitie-filosofisch debat dat gaat over de essentie van menselijke cognitie intelligentie, en of die essentie te recreëren, dan wel te emuleren, is in kunstmatige systemen. Een deel van de wetenschappers en futuristen zien in de huidige snelle ontwikkelingen tekenen dat 'singularity' (het moment waarop technologie menselijke intelligentie

[13] Over het ontstaan van de term 'Artifical Narrow Intelligence' en 'Artifical General Intelligence': Ben Goertzel, "Who Coined the Term AGI," https://goertzel.org/who-coined-the-term-agi/.

[14] Andreas Kaplan, Michael Haenlein, "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons* 62, no. 1 (januari-februari 2019), https://www.sciencedirect.com/science/article/pii/S0007681318301393

[15] Steve Crowe, "Humanoid robots: Five to watch in 2019," *The Robot Report* 29 januari 2019, https://www.therobotreport.com/humanoid-robots-watch-2019/.

bereikt) halverwege de 21e eeuw bereikt zal worden[16]. Andere zien het als mythe [17], mythisch doel [18] of iets dat nog vele generaties weg is [19].

'*Artificial Super Intelligence*' (ASI) wordt gezien als de overtreffende trap van AGI. Als een systeem als 'ASI' geclassificeerd wordt, dan zou het op alle fronten menselijke cognitieve capaciteiten overtreffen, en zou het opereren op een manier die ook niet meer goed door mensen te interpreteren is. Net zoals AGI wordt ASI voornamelijk gezien als een hypothetisch construct aangezien AI-technologie en toepassingen nog lang niet zo vergevorderd zijn dat er over praktische ASI gesproken kan worden. Het concept 'Superintelligence' (ASI) wordt op dit moment vooral gebruikt om ethische, juridische en ideologische vraagstukken over de rol van systemen in de toekomst te voeden.

De onderverdeling in ANI, AGI en ASI is belangrijk om in beschouwing te nemen, omdat het tekenend is voor de ambities van degenen die AI ontwikkelen of inzetten. Een systeem met AGI oogmerk heeft potentieel een heel ander impactprofiel dan een systeem dat als ANI getypeerd is. De classificatie 'AGI' (en 'ASI') impliceert dat het systeem in kwestie blijft leren, blijft ontwikkelen, zich blijft aanpassen aan z'n omgeving, zodat het steeds beter in staat wordt om taken uit te voeren – op een zelfde manier waarop de menselijke geest zich ontwikkelt en aanpast aan nieuwe omstandigheden.

Daadwerkelijke AGI is nog ver weg, maar veel AI-toepassingen worden ontwikkelend in die gedachte. Denk bijvoorbeeld aan grote data verwerkende bedrijven die hun systemen steeds meer laten leren en afleiden over gebruikers, of de ontwikkeling van militaire autonome platformen die steeds meer zelfstandig kunnen opereren, zoals drones of robots in het landoptreden.[20]

---

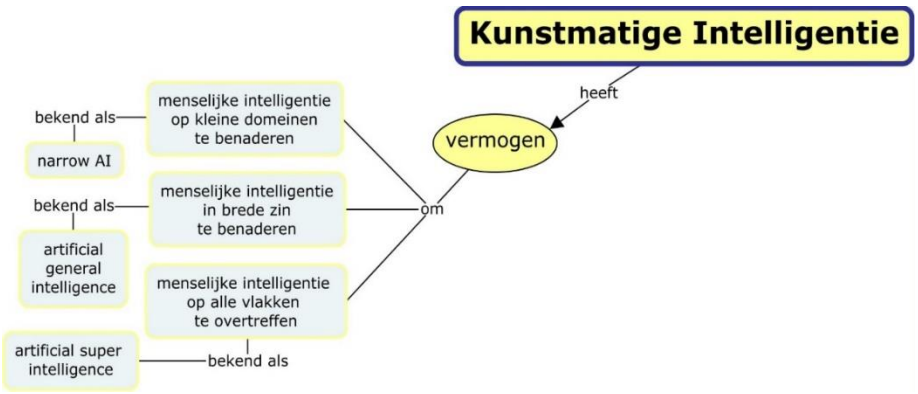
[16] Robert Friday, "A World Run By Intelligent Machines: How Close Are We?," *Forbes* 2 januari 2020, https://www.forbes.com/sites/forbestechcouncil/2020/01/02/a-world-run-by-intelligent-machines-how-close-are-we/#447300594088.

[17] Rodney Brooks, "The Seven Deadly Sins of AI," *MIT Technology Review* 6 oktober 2017, https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/

[18] Kevin Kelly, "The AI Cargo Cult. The Myth of a Superhuman AI," *Wired* 25 april 2017. https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/

[19] James Vincent, "This is when AI's top researchers think artificial general intelligence will be achieved," *The Verge* 27 november 2018, https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book.

[20] Tannya D. Jajal, "Distinguishing between Narrow AI, General AI and Super AI," *Medium* 21 mei 2018, https://medium.com/@tjajal/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22.



Figuur 4: Cmap Kunstmatige Intelligentie – Vermogen

3.2.2 Hoofdconcept: Methoden, modellen en algoritmes

Kunstmatige Intelligentie is een overkoepelende term voor een grote familie van methodes, modellen en algoritmes. Deze familie is sterk gegroeid sinds de term 'Kunstmatige Intelligentie' ontstond in wetenschappelijk debat halverwege de jaren '50.[21] Veel van deze algoritmes zijn al in de beginjaren van AI bedacht, maar werden pas later praktisch toepasbaar door de ontwikkeling van computerrekenvermogen, netwerken en dataopslag.

Van oudsher zijn er verschillende archetypen benaderingen van kunstmatige intelligentie. De verschillende perspectieven hebben vooral te maken met hoe kennis wordt gerepresenteerd in het systeem en hoe het leerproces verloopt. De Cmap (Figuur 5) toont de belangrijkste onderverdelingen en varianten die uit de verkenning en de expertsessies voortgekomen zijn.

Vorm van leren

De vorm van leren is een belangrijke discriminator tussen verschillende modelfamilies. Er zijn drie hoofdtypes: *supervised learning*, *reinforcement learning* en *unsupervised learning*. Dit onderscheid gaat over de wijze waarop een AI-systeem getraind wordt om taken uit te voeren. Als een systeem 'supervised' getraind wordt dan biedt een ontwerper het systeem data aan en beoordeelt of het systeem de aangeboden data goed geïnterneerd heeft. Denk hierbij bijvoorbeeld aan het aanbieden van beelden van verkeersborden aan een besturingssysteem voor autonome voertuigen, of het aanbieden van gebruikersprofielen aan een beoordelingssysteem dat gebruikers moet categoriseren.

In een 'unsupervised' variant traint het systeem zichzelf. Het systeem neemt daarbij data op en beoordeelt zelf of zijn gedrag succesvol is. Dit kan vergeleken worden met menselijk 'uitproberen'. Een goed voorbeeld hiervan zijn bijvoorbeeld big data-applicaties waar systemen nieuwe diagnoses of inzichten ontdekken. Ontwerpers trainen deze inzichten niet, maar bouwen een geschikte context

[21] John McCarthy, Marvin L. Minsky, Nathaniel Rochester en Claude E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," reproduced in *AI Magazine* 2, no.4, 2006, p.12.

waarbinnen systemen kunnen leren (voorbereiden van data, definiëren van functioneringsindicatoren, etc.).

Bij reinforcement learning zoekt het systeem zelf naar een oplossing voor een vraag, en krijgt een beloning van de ontwerper als deze oplossing de vraag goed beantwoordt. De beloning is hoger naarmate de oplossing de vraag beter beantwoordt. De ontwerper beoordeelt hierbij dus niet *hoe* het systeem tot de oplossing komt, maar hoe dicht de oplossing bij de verwachte uitkomst komt. Het systeem probeert de beloning te maximaliseren, en zal dus oplossingsrichtingen die een positieve beloning opleveren, verder verdiepen en versterken ('*reinforce*'). Deze aanpak is interessant bij vraagstukken waar geen vooropgezette oplossingen zijn, en de zoekruimte groot is, zoals bij navigatie van autonome platformen, in de besturing van industriële systemen, de werking van aanbevelingssystemen in digitale winkels en het trainen van gedrag van actoren in simulatieomgevingen.

Vorm van kennis representeren

Een andere belangrijke dimensie waarop de AI-methodes onderling kunnen verschillen is de vorm waarin kennis gerepresenteerd is binnen het systeem. Er zijn twee fundamentele stromingen: expliciete kennisrepresentatie en impliciete kennisrepresentatie. Dit onderscheid geeft aan hoe kennis (concepten, regels, kennis) gerepresenteerd is in het systeem. We spreken van expliciete (of symbolische) kennisrepresentatie als er gebruik gemaakt wordt van voor ons herkenbare concepten, veelal woorden in taal. Denk hierbij bijvoorbeeld aan traditionele kennissystemen (*rule-based systems*) die werken met als-dan regels, en waarbij begrippen semantisch vastgelegd zijn. Andere voorbeelden van methodes die gebruik maken van expliciete kennisrepresentatie zijn *case-based reasoning*, *semantic reasoning* en *decision trees*. In een case-based reasoning systeem vergelijkt het systeem een aangeboden oplossing met opgeslagen 'cases': vastgelegde verslagen van gebeurtenissen of handelswijzen. Het systeem beoordeelt volgens een redeneerproces de mate waarin de aangeboden case overeenkomt met een opgeslagen case, en geeft daarmee een mate van herkenning. Semantic reasoning en decision trees methodes doen iets gelijkwaardigs: ze vergelijken een aangeboden observatie met bekende beschrijvingen van een domein of een proces, en geven een beoordeling terug. Zo kan bijvoorbeeld via een decision tree een AI-systeem het type dier herkennen op basis van kenmerken, of een medisch systeem een diagnose stellen op basis van symptomen. Een semantic reasoning systeem maakt gebruik van semantische netwerken waarin beschreven is hoe concepten aan elkaar gelinkt zijn, zoals dat een 'fiets' een subtype is van het concept 'voertuig', en dat het woord 'explosie' nauw verwant is met het woord 'ontploffing'. Door te redeneren over concepten, kan een semantisch netwerk bijvoorbeeld vaststellen dat twee publicaties inhoudelijk dicht bij elkaar liggen, of dat een tekst in een bepaalde categorie valt.

Fuzzy logic en *Bayesian logic* zijn statistische methodes die werken op basis van verwachtingen en onzekerheden. Ze maken gebruik van bekende afhankelijkheden tussen concepten, en berekenen de kans dat een observatie behoort tot een bekend patroon.

We spreken van impliciete (of subsymbolische) kennisrepresentatie als de kennis in een systeem niet vastgelegd is in voor mensen herkenbare taal. Denk hierbij aan getallen, codes of andere tekens. Neurale netwerken zijn exemplarisch hiervan,

omdat de kennis vastgelegd is in matrices van getallen (gewichten) en als zodanig niet direct te interpreteren is. Onder deze noemer vallen ook veel biologisch geïnspireerde technieken zoals genetische algoritmes die oplossingen 'combineren' en laten evolueren totdat een oplossing gevonden wordt die aan succescriteria voldoet. *Particle swarm optimization* maakt gebruik van 'zwerm' principes waarbij oplossingen 'swarmen' naar gebieden van succesvolle oplossingen.

Support vector machines en *Artificial Neural Networks* zijn voorbeelden van patroonherkenningstechnieken die gebruik maken van '*data fitting*': ze proberen hun interne patronen zo goed mogelijk in te stellen om aangeboden patronen te spiegelen. Support vector machines maken hierbij gebruik van vectoren (punten in een oplossingsruimte) om een oplossing te representeren. Neurale netwerken maken gebruik van 'neuron' en de gewichten tussen neuronen om een oplossing te representeren. Beide technieken worden breed toegepast, en steeds vaker in combinatie.[22]

Momenteel worden ook veel hybride varianten van deze archetypes ontwikkeld, zoals bijvoorbeeld *Generative Adversarial Neural Networks (GAN)*, waarbij twee neurale netwerken *elkaar* trainen; één neuraal netwerk is getraind met data, de ander genereert oplossingen. Het getrainde netwerk geeft een beloning als de gegeneerde oplossing op getrainde data lijkt. Het ongetrainde netwerk probeert de beloningen te maximaliseren en leert zo steeds beter de getrainde data te benaderen. Dit proces leidt tot oplossingen die lijken op de getrainde set, maar niet identiek, en geven daarmee een mate van creativiteit. Dit principe is bijvoorbeeld al uitgebuit in verschillende '*art generation*' demonstraties[23] waarin GANs 'schilderijen' creëren in de geest van bekende schilders[24] [25]en muziek genereren.[26]

Systeem scope (Embodiment). Een andere vaak gebezigde dimensie is de opdeling in '*embodied'* en '*non-embodied*' systemen. De onderliggende vraag bij deze indeling is of het AI-systeem een fysieke vorm heeft of een virtuele. Een fysiek AI-systeem is bijvoorbeeld een robot of een drone. Dit soort systemen moet kunnen observeren en opereren in de fysieke wereld, en heeft derhalve sensoren (camera's, meetinstrumenten, locatiesystemen, etc.) en actuatoren (armen, wielen, wapens, bewegingsinstrumenten). Een virtueel, non-embodied systeem is een systeem dat alleen digitaal bestaat en heeft dus niet de beschikking over fysieke koppelingen met de buitenwereld (bijvoorbeeld sensoren, actuatoren, interactie-

[22] Manikandan Jeeva, "The Scuffle Between Two Algorithms - Neural Network vs. Support Vector Machine," *Medium* september 2018, https://medium.com/analytics-vidhya/the-scuffle-between-two-algorithms-neural-network-vs-support-vector-machine-16abe0eb4181.

[23] "Artificial Art: How GANs are making machines creative. Examining the creative potential of machines," *Medium* 23 september 2019, https://heartbeat.fritz.ai/artificial-art-how-gans-are-making-machines-creative-b99105627198.

[24] James Vincent, "How Three French Students Used Borrowed Code To Put The First Ai Portrait In Christie's," *The Verge* (23 oktober 2018), https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans;

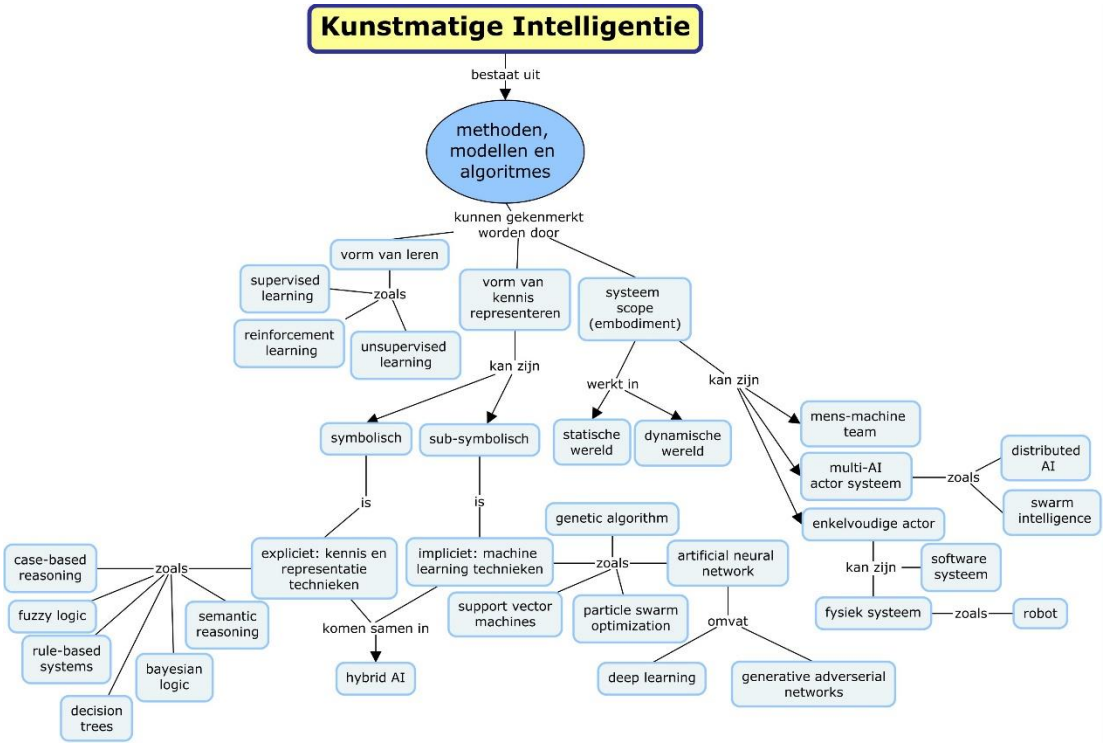
[25] Kenny Jones, "GANGogh: Creating Art with GANs," *Medium* 18 juni 2017, https://towardsdatascience.com/gangogh-creating-art-with-gans-8d087d8f74a1

[26] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, Yi-Hsuan Yang, "MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment," november 2017, https://arxiv.org/abs/1709.06298.

middelen). Dit onderscheid is van belang om begrip te krijgen van de positionering van een AI-systeem: bevindt deze zich in de fysieke wereld, de digitale wereld, of beide? Deze werelden hebben eigen normen en regels, en daarom is begrip van de vorm van een AI-systeem van belang.

Een andere belangrijke 'embodiment' typering is hoeveel actoren tot het AI-systeem behoren. Het systeem kan bestaan uit een enkelvoudig AI-systeem, zoals een enkele robot, of een enkel softwaresysteem, of bestaan uit meerdere systemen, een zogenaamd multi-AI-systeem. Hierbij kan bijvoorbeeld gedacht worden aan een zwerm drones ('swarming intelligence') die onderling communiceren en coördineren, of een verzameling software *agents* die gezamenlijk verschillende taken uitvoeren. Een andere variant is een mens-machine team, waarin een menselijke gebruiker en een AI-systeem nauw samenwerken, zoals het geval is in voertuigen of *command and control* situaties.

Er zijn vele andere onderverdelingen mogelijk, en er verschijnen ook nieuwe algoritmevariaties. De nieuwigheid zit veelal in de combinatie van methodes, en het integreren van andere technologieën met AI, zoals *data science* technieken, *distributed network* modellen en modellerings- en simulatieomgevingen.



Figuur 5: Cmap Kunstmatige Intelligentie - methoden, modellen en algoritmes

3.2.3 Hoofdconcept: Functies

Kunstmatige Intelligentie gaat in beginsel over het vermogen van systemen om menselijke cognitieve functies uit te kunnen voeren, of aan te vullen. Veel van de typische toepassingen van AI-algoritmes vervullen daarom systeemtaken die we traditioneel 'cognitief' noemen: redeneren, leren, herkennen, zien, improviseren,

beargumenteren, en dergelijke. *Deep learning* algoritmes kunnen bijvoorbeeld worden ingezet ten behoeve van patroonherkenning in grote hoeveelheden data. Een systeem leert zichzelf dan om patronen te herkennen in data, zoals tekst of afbeeldingen. Deze functies kunnen op verschillende manieren geïmplementeerd worden, met verschillende soorten AI-algoritmes.

AI-algoritmes worden veelal ingezet om een bestaande systeemfunctie te versterken of te versnellen, en 'AI' is derhalve vaak meer een aanvulling dan een vernieuwing voor veel systemen. De meest in het oog springende toepassingen zijn doorontwikkelingen van systemen die al bestonden, maar nu veel beter zijn geworden, en daarmee meer impact hebben. Denk aan autonome platformen, online winkelsystemen en surveillance systemen; deze zijn de afgelopen jaren veel capabeler geworden door de toename van rekenkracht én de inzet van AI-algoritmes.

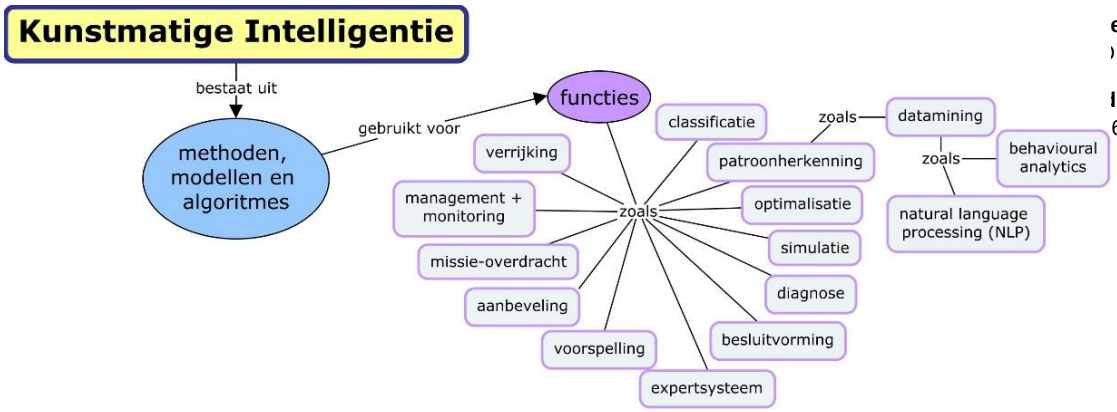
Veel van de genoemde functies (Figuur 7) zijn ook nauw aan elkaar gelinkt. Patroonherkenning is als functie cruciaal voor het kunnen voorspellen, bijvoorbeeld in een medische diagnose context. Classificatie is als functie essentieel voor data-analyse, en dus voor bijvoorbeeld patroonherkenning of het genereren van aanbevelingen. AI wordt ook veel toegepast als drijvende kracht achter *datamining* applicaties; het identificeren van relevante patronen in grote datasets, zoals big-data vraagstukken in commerciële en publieke sectoren. Het verwerken en interpreteren van natuurlijk taal ('*natural language processing*') is ook een voorbeeld van datamining waarbij veel achtergrondkennis en dataprocessing aan de pas komt. Op basis van dit soort functies leren machines steeds beter menselijk taalgebruik te begrijpen en menselijk te communiceren. Deze ontwikkelingen gaan snel en het wordt steeds moeilijker om kunstmatige van natuurlijk gegeneerde taal te onderscheiden. Zie bijvoorbeeld de ontwikkelingen rondom Google Duplex[27] en de huidige generatie chatbots die bijna vlekkeloos informeel gesproken ontleden.

'*Behavioural analytics*' is een ander voorbeeld van datamining waarbij op basis van geobserveerd gedrag (bijvoorbeeld online gedrag van personen en groepen, of bewegingen van verdachte actoren) een gedragsprofiel opgesteld wordt en een inschatting gemaakt kan worden van voorkeuren en toekomstig gedrag. Dit vormt bijvoorbeeld de basis van veel inlichtingenwerk en '*predictive policing*'.

Het is belangrijk om de functie van de toepassing en de implementatie te onderscheiden en te begrijpen waar een AI-technologie precies aan bijdraagt in een systeemcontext. De functie die een AI-algoritme speelt in een systeem heeft namelijk invloed op het impactprofiel.

De functies die genoemd worden in de Cmap vormen een goede vertegenwoordiging van typische functies die AI-algoritmes vervullen in systemen, maar zouden kunnen ook op andere manieren geordend worden, bijvoorbeeld op basis van lagere of hogere cognitieve functies, of op basis van output.

[27] Brian X. Chen en Cade Metz, "Google's Duplex Uses A.I. to Mimic Humans (Sometimes)," *The New York Times* 22 mei 2019, https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html.



Figuur 6: Cmap Kunstmatige Intelligentie - Functies

3.2.4 Hoofdconcepten: Toepassingen & toepassingsfuncties en sectoren (inclusief risico's en dreigingen)

De hoofdconcepten toepassingen & toepassingsfuncties (groen) en sectoren (roze) worden in dit deel gezamenlijk besproken.

De verschillende functies zoals ze hierboven staan beschreven (in paragraaf 3.2.3) worden concreet toegepast in toepassingen ten behoeve van bepaalde taken (toepassingsfuncties) in bepaalde sectoren. Ter illustratie is een aantal voorbeelden toegevoegd aan deze Cmap bij een aantal sectoren (zie Figuur 8), maar de mogelijkheden zijn natuurlijk bijna oneindig. Het concrete toepassingsvoorbeeld van onbemande autonome wapensystemen binnen de veiligheidssector heeft bijvoorbeeld als één van de taken om aan *sensing* te doen (concreet: op zoek gaan naar het target)

Nagenoeg alle sectoren in de samenleving hebben al te maken met op AI-gebaseerde systemen of maken verandering door als gevolg van AI-gedreven toepassingen. AI-technologie zit dusdanig diep verweven in de samenleving dat het meer de vraag wordt hoe je AI herkent. Steeds minder producten afficheren zich expliciet als AI-gedreven en er komen steeds meer andere metaforen voor AI in zwang, zoals 'zelflerend', 'autonoom', 'adaptief' en '*smart*'. Producten die zich als zodanig kenmerken maken meestal veelvuldig gebruik van AI-algoritmes.

De belangrijkste reden om te kijken naar sectoren is om te identificeren waar AI-technologieën significante veranderingen teweeg gaan brengen, en waar potentieel een impact op nationale veiligheid gaat ontstaan. In de Cmap staat een beknopte opsomming van sectoren die tijdens de sessies genoemd zijn als voorbeeld waar AI-technologie wordt toegepast en waar belangrijke veranderingen verwacht worden.

De snelheid waarmee sectoren een AI-gedreven verandering doormaken heeft niet alleen te maken met de waarde die AI-toepassingen brengen. Ook regulatie, risicoafwegingen en randvoorwaarden hebben grote invloed. De toepassing van AI-algoritmes in de commerciële dienstensector gaan bijvoorbeeld harder dan de militaire of justitiële sector door beperktere regulatie. Daarnaast is ook niet in alle sectoren de meerwaarde van AI-technologie even voor de hand liggend. Vooral

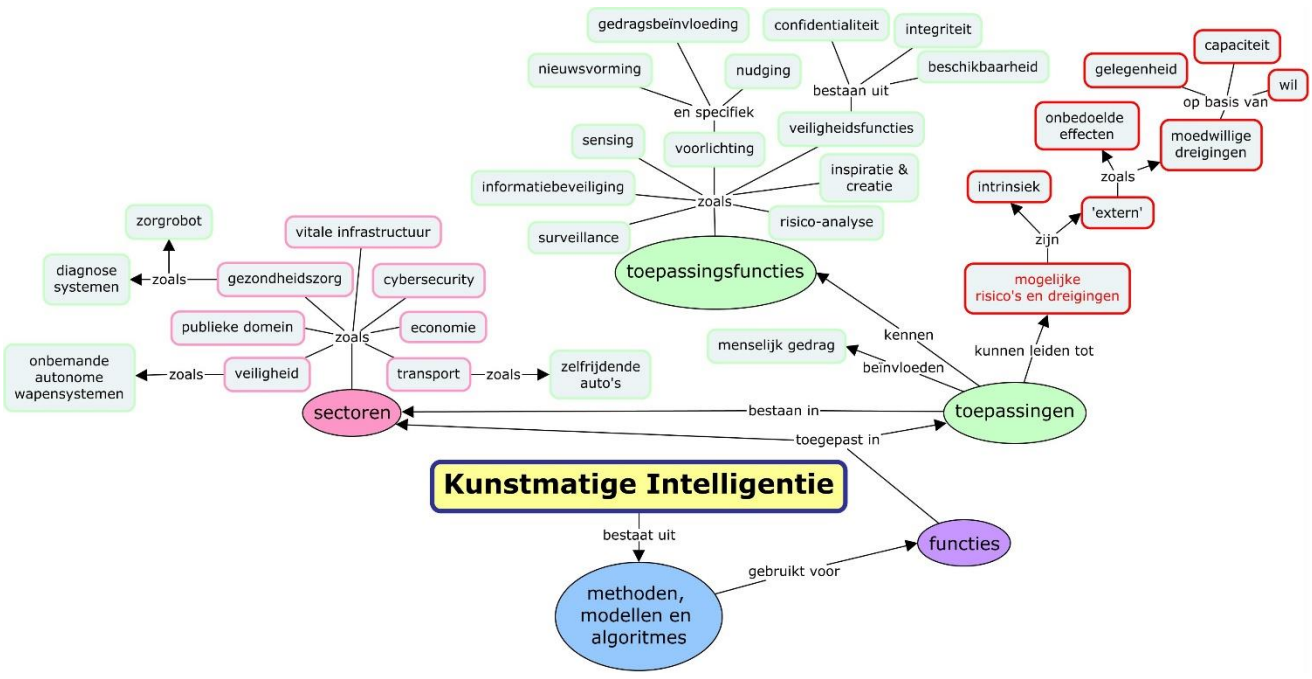
sectoren met complexe planningsuitdagingen of grote datasets zien veel ontwikkelingen, zoals de logistieke sector of de *intelligence communities*.

Een sector waar AI-technologie grote impact gaat hebben is de cybersecurity sector. AI-technologie zal steeds meer een factor van belang worden in zowel het dreigingslandschap als in het tegengaan van digitale bedreigingen.

Door het gebruik van AI-technologie in digitale aanvallen en malware zullen deze steeds complexer en grootschaliger worden, en lastiger tegen te houden met conventionele cybersecurity strategieën. Cybersecurity-producten zullen met behulp van AI-technologie zelflerend, zelfsturend en samenwerkend gemaakt moeten worden om deze dreigingen het hoofd te bieden. De eerste indicaties van deze AI-tegen-AI strijd zijn nu al zichtbaar en zullen de cybersecurity sterk beïnvloeden. [28]

3.2.4.1 Mogelijke risico's en dreigingen

Gedurende dit project is veel gesproken over risico's en dreigingen rondom het gebruik van AI-technologie. Deze vraag is minder makkelijk te beantwoorden dan het lijkt. AI-technologie is op zichzelf noch risico noch dreiging. Toegepast in een product kan het een risico gaan vormen door het typisch zelfsturende, zelflerende karakter van veel AI-gebaseerde producten. Maar wat voor soort risico is dat dan, en hoe beschrijven we de zwaarte en implicatie daarvan?



Figuur 7: Cmap Kunstmatige Intelligentie – sectoren, toepassingen & toepassingsfuncties

[28] Louk Faesen en Erik Frinking, *Understanding the Strategic and Technical Significance of Technology for Security Implications of AI and Machine Learning for Cybersecurity* (The Hague Centre for Strategic Studies, augustus 2019), https://hcss.nl/report/understanding-strategic-and-technical-significance-technology-security-implications-ai-and.

Er speelt nog een ander fenomeen mee als we het hebben over de risico's van AI-technologie: de wereld staat niet stil. Er wordt veel tijd en geld geïnvesteerd in het versterken van AI-technologieën, en het aanpakken van bekende uitdagingen zoals data bias en beperkte transparantie. Veel van de kenmerkende risico's die met AI-technologie gemoeid zijn, zouden mogelijk weggenomen kunnen zijn als je 5 of 10 jaar de toekomst in kijkt. Daarnaast verandert de maatschappij zelf ook. De ethische normen die we nu proberen op te stellen rondom AI-technologie zijn wellicht niet meer relevant in de (nabije) toekomt door veranderende perceptie jegens AI-toepassingen en/of anders ingerichte wetgeving. Daarbij komt nog dat op termijn de maatschappij zelf gaat veranderen en er dus mogelijk heel andere risico's kunnen gaan gelden.

3.2.5 Hoofdconcept: Actoren

Bij de ontwikkeling en het gebruik van AI-toepassingen spelen veel verschillende actoren een rol, zowel nationaal als internationaal (Figuur 9). Producenten van AI-systemen hebben een grote rol, maar ook gebruikers van deze systemen die de vraag naar AI-toepassingen mede vormgeven. Aan de ontwikkelkant speelt de wetenschap een rol met de doorontwikkeling van bestaande methodes en beschouwingen rondom ontwerp en inzet van AI-technologie. Ook technologie-bedrijven zoals Google, Facebook en Microsoft investeren significant in vroeg onderzoek en doen dat veelal in open verbanden.

Daarnaast zijn er veel actoren die een minder zichtbare, doch significante rol hebben in de introductie van AI in onze maatschappij. AI-technologieën kunnen niet functioneren zonder data. De snelle ontwikkeling van AI-gedreven toepassingen maakt dat er ook een data-economie ontstaan is rondom het creëren, analyseren, compileren en verrijken van allerhande data.

Naarmate AI meer gemeengoed technologie wordt, gaan meer actoren betrokken worden. Denk bijvoorbeeld aan de activiteiten die zich nu afspelen rondom regulatie en verzekeringen, en de grote vraagstukken rondom governance, ethiek en wetgeving. De vraag rijst dan ook of het nog zinnig is om over 'actoren in het AI-speelveld' te spreken aangezien nagenoeg elke actor in een maatschappij, van burger tot bedrijf, van producten tot overheid met manifestaties van AI te maken heeft, dan wel krijgt.

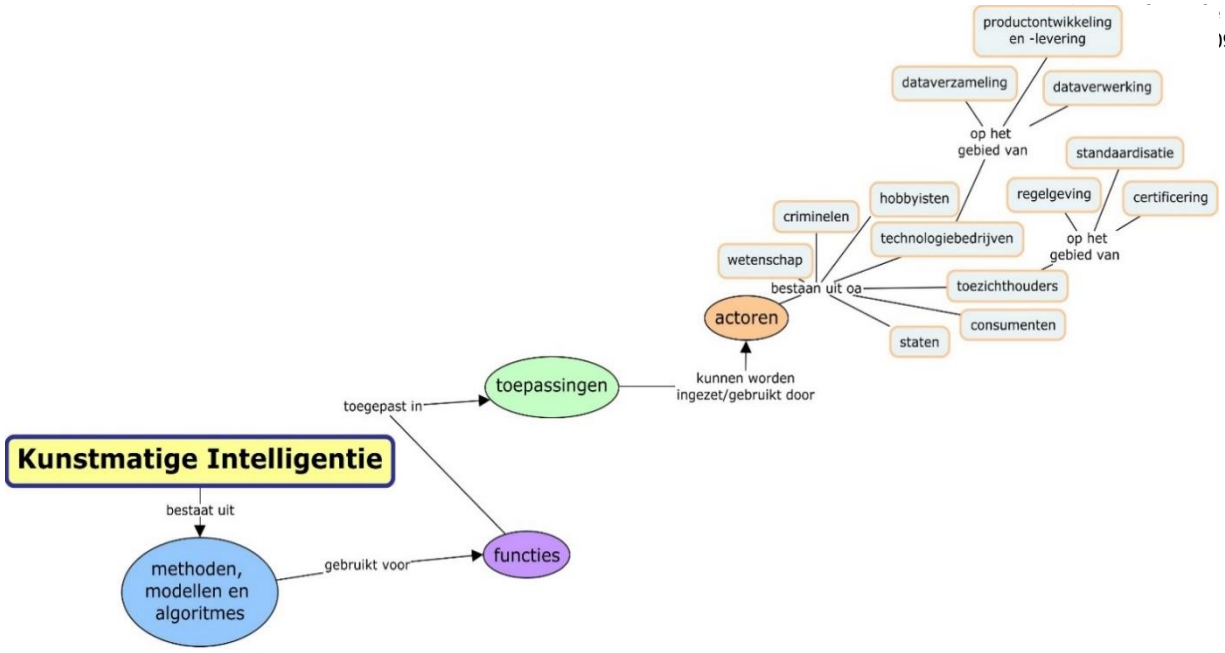
In de Cmap is een voorbeeldcollectie van actoren weergegeven in een simpele onderverdeling zoals die uit de expertsessies voortkwam. Deze set zou ook weergegeven kunnen worden in de vorm van complexen: netwerken van actoren die gezamenlijk een ketensysteem vormen. Hieronder staat een aantal voorbeelden van complexen die verband houden met de introductie van AI in de samenleving:

- Industrieel complex: Het collectief van soft- en hardware producenten, toeleveranciers, ontwerpers en andere actoren die betrokken zijn bij de productie en levering van AI-gedreven producten.
- Beleidscomplex: Het collectief van actoren die beleidsmatig sturing geven aan de introductie van AI-systemen in de maatschappij, het stellen van (juridische) kaders, het toezien op handhaving, en het stellen van prioriteit in relatie tot veiligheid, economische ontwikkeling of sociale zaken.
- Innovatiecomplex: De keten van partijen die vroege, fundamentele, en toegepaste wetenschap uitvoeren en het innovatieproces rondom de

toepassing van AI-technologieën vormgeven. Denk hierbij aan universiteiten, kennisinstituten, innovatielabs en *research and development* afdelingen van technologiebedrijven, maar ook open source groeperingen.
- Maatschappelijk complex: Het collectief van actoren die vanuit maatschappelijk perspectief de introductie van AI beïnvloeden, zoals belangenverenigingen, de media, zorginstellingen, activistische organisaties, ethische en privacy advocaten.
- Financieel complex: Het collectief van actoren die een rol spelen in de financiering van bijvoorbeeld AI-onderzoek. Denk hierbij aan de bankensector, investeerders, overheden, particuliere fondsen, verzekeraars en *crowdfunding* initiatieven.

Er zijn er nog veel meer complexen te definiëren. Dit is niet meer dan logisch gezien de maatschappelijke transformatie die AI-technologie voortbrengt. Actoren zullen daarom ook vaak een rol spelen in meerdere complexen. Tevens kunnen deze complexen zowel nationale als internationale actoren omvatten.

Een wat minder duidelijk herkenbaar complex is het ‘kwaadwillenden complex’. AI-technologie brengt niet alleen positieve maatschappelijke ontwikkelingen met zich mee maar geeft ook kwaadwillende actoren nieuwe middelen en nieuwe mogelijkheden. Hierbij kan gekeken worden naar bijvoorbeeld statelijke actoren, of criminele organisaties, maar ook naar alle partijen die ertoe bijdragen dat deze partijen AI-gebaseerde capaciteiten verkrijgen. Hieronder kunnen actoren vallen die niet van zichzelf kwaadwillend zijn, maar die wel misbruik van AI-technologie in de hand werken. Denk bijvoorbeeld aan datacentra, netwerkproviders, open source collectieven die AI-algoritmes beschikbaar stellen en producenten die producten leveren die misbruikt kunnen worden.



Figuur 8: Cmap Kunstmatige Intelligentie - Actoren

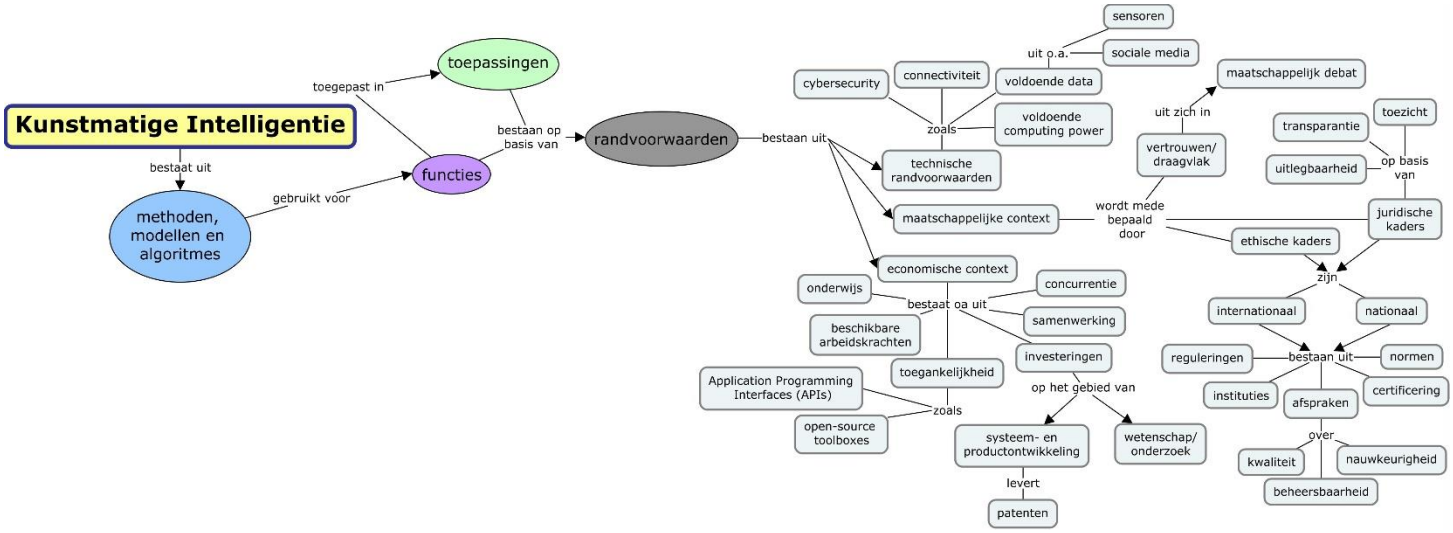
3.2.6 Hoofdconcept: Randvoorwaarden

AI-toepassingen en -systemen bestaan niet in een vacuüm, maar hebben te maken met context. Deze context representeert alle zaken die zich afspelen rondom een systeemtoepassing en alle zaken die het mogelijk maken dat een systeem zijn functie uitvoert. De context omvat derhalve de eigenschappen van de omgeving waarin het systeem werkt. Dit hebben we gevat onder het hoofdconcept randvoorwaarden en opgedeeld in technische randvoorwaarden, maatschappelijke context en economische context (Figuur 10).

Onder technische randvoorwaarden vallen noodzakelijke infrastructurele elementen om een AI-systeem goed te kunnen laten functioneren. De meeste AI-toepassingen zijn data intensief, en vergen grote hoeveelheden data om getraind te worden, of om te opereren. Hiervoor dienen de juiste bronnen aanwezig te zijn, evenals een geschikte data infrastructuur om de toevoer van data te organiseren. Naast data is rekenkracht (*computing power*) een factor van belang. Veel AI-algoritmes zijn rekenintensief, bijvoorbeeld complexe neurale netwerken. Als de rekenkracht niet op het productplatform aanwezig kan zijn dan moeten daar voorzieningen voor getroffen worden, bijvoorbeeld door berekeningen op een andere locatie uit te voeren (*cloud computing*). Ook (cyber)security is een belangrijke randvoorwaarde: het voorkomen dat AI-toepassingen door kwaadwillenden beïnvloed worden. Kwaadwillenden kunnen bijvoorbeeld de datasets beïnvloeden waarop AI-algoritmes werken (‘*data poisoning*’), de algoritmes beïnvloeden zodat ze zich anders gedragen, of daadwerkelijk uitschakelen. Naarmate AI-algoritmes in steeds meer vitale maatschappelijke systemen en processen een rol speelt, wordt het belang van deugdelijke fysieke en digitale bescherming steeds groter.

AI-toepassingen bestaan ook in een maatschappelijke context. Dit beslaat onder meer juridische kaders en ethische kaders (zowel nationaal als internationaal). Ook de mate waarin de samenleving vertrouwen heeft in AI-systemen heeft invloed op de manier waarop AI-systemen worden toegepast. Met enige regelmaat komen er berichten over AI-toepassingen die verkeerde beslissingen nemen als gevolg van databias, onvolledige training of misfunctionering. Dit soort berichten maakt dat er een scepsis kan ontstaan jegens AI-toepassingen.

Een derde type randvoorwaarde is de economische context. AI als technologie kan alleen worden doorontwikkeld als er voldoende investeringen gedaan worden in research & development en het innovatie- en implementatieproces. Daarnaast moeten er rondom de AI-toepassing ook genoeg flankerende structuren zijn, zoals bedrijven die de noodzakelijke data en rekenkracht leveren, of bedrijven die aan de voorkant uitkomsten verder verwerken.



Figuur 9: Cmap Kunstmatige Intelligentie - Randvoorwaarden

3.3 Gebruik van de concept map

De concept map ‘verkenning Kunstmatige Intelligentie in de context van nationale veiligheid’ is voornamelijk bedoeld om een gezamenlijk denkkader te scheppen op basis waarvan een inhoudelijke discussie kan plaatsvinden. Zowel voor analisten als voor beleidsmedewerkers biedt een dergelijke kennisrepresentatie een hulpmiddel om het speelveld te kunnen blijven overzien en een concept map faciliteert op die manier een brede blik en helpt een tunnelvisie te voorkomen.

Concreet kan een dergelijke Cmap worden gebruikt door analisten en beleidsmedewerkers als een ‘navigator’ voor het identificeren van mogelijke AI-gerelateerde risico’s en dreigingen. Verschillende combinaties van concepten op de kaart leveren namelijk een ‘narratief’ (een mini-scenario) van een dreiging of risico op. Inhoudelijk zal hierop in het volgende hoofdstuk verder worden ingegaan.

Een andere manier van toepassing van de concept map is het simpelweg inspireren van analisten en beleidsmedewerkers. Het inspireren geldt in belangrijke

mate ook tijdens het maken van de concept map zelf in een expertsessie, waar verschillende meningen samen gebracht worden in de map. De expertinput verrijkt niet alleen de concept map, maar ook de kennis van de experts zelf. Door experts vanuit verschillende achtergronden en met verschillende invalshoeken samen te laten komen tot een gedeeld denkkader (de concept map), worden de experts zelf namelijk ook blootgesteld aan perspectieven die zijzelf wellicht tot dan toe in mindere mate hadden onderkend.

Ook kan de concept map worden gebruikt als een spiegel voor beleidsmakers. De thema's op de kaart worden gekoppeld aan staand beleid, waardoor eventuele gaten in beleid kunnen worden geïdentificeerd. Dat wil niet zeggen dat een dergelijke hiaat in beleid altijd opgevuld moet worden met nieuw beleid, maar het is wel relevant te onderkennen dat er een dergelijk hiaat (of misschien blinde vlek) bestaat. Naast het identificeren van gaten in beleid kan een concept map gebruikt worden als 'check' op voorgenomen beleid. Zijn de verschillende afhankelijkheden binnen een technologiegebied als AI voldoende onderkend (en bekend) in het voorgenomen beleid?

Bovendien helpt een concept map en de gehele exercitie om tot een concept map te komen voor begripsvorming bij beleidsmedewerkers in een relatief korte tijd over een complex onderwerp als AI-technologie. Dit faciliteert vervolgens het vraagarticulatieproces richting analisten of kennisinstituten indien er behoefte is aan verdere (impact)analyses.

4 Eerste indicaties mogelijke risico's en dreigingen

4.1 Impact op de nationale veiligheid: risico's en dreigingen

Bij een analyse van de impact van technologie op nationale veiligheid zijn beide zijden van dezelfde munt van belang: de kansenkant en de dreigingskant. In dit project is in afstemming met de opdrachtgevers gekozen voor een focus op de dreigingskant van de impact van AI op nationale veiligheid.[29]

Maar wat is dan precies een risico of dreiging van AI? In generieke zin kan die vraag niet beantwoord worden, zoals eerder opgemerkt. Als we AI puur beschouwen als technologie, dan valt daar geen risico aan toe te kennen, net zo min als aan een 'database' of een 'netwerk' een inherent risico verbonden is. Pas zodra technologie wordt toegepast of gevuld met data en diensten ontstaan risico's, zoals uitval, corruptie en misbruik.

In het project hanteren wij het onderscheid tussen een risico en een dreiging, zoals dat ook wordt gehanteerd in de Nationale Veiligheid Strategie en binnen het Analistennetwerk Nationale Veiligheid.

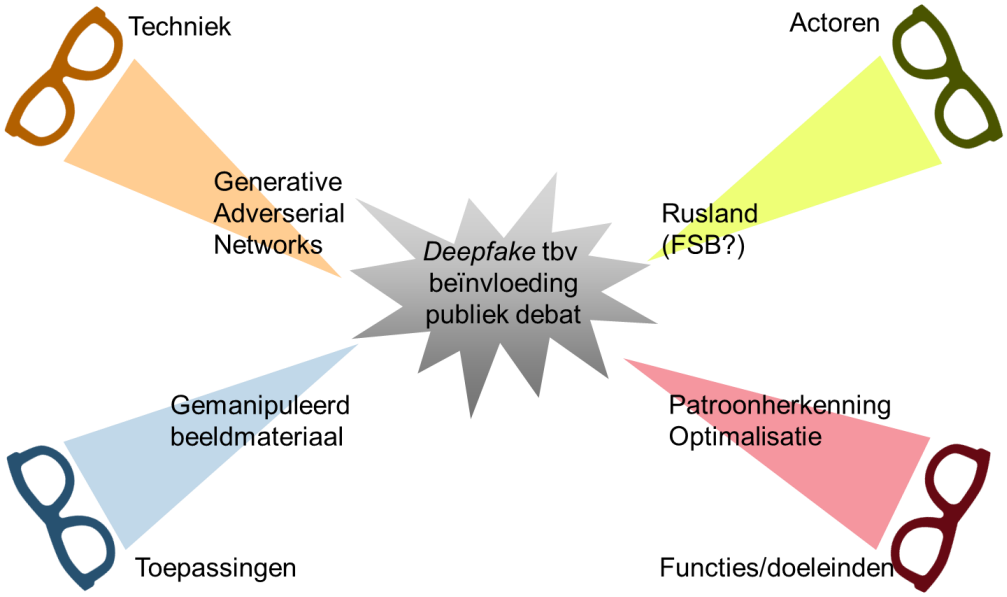
- Een risico is het samenspel van de waarschijnlijkheid dat een incident zich voordoet en de impact die dat kan hebben;
- Een dreiging duidt op aanwezigheid, concreetheid en acuutheid van gevaar (soms nog als een stip aan de horizon die via *early warning* onderkend moet worden).

Een risico ligt dus kort gezegd verder in de toekomst (het is een potentieel gevaar) en een dreiging is een aanwezig gevaar in het nu.[30]

4.2 Eerste verkenning dreigingen en risico's

Tijdens de eerste expertsessie met technisch-inhoudelijke experts op het gebied van AI is naast het bespreken van de concept map ook aandacht geschonken aan mogelijke risico's en dreigingen. Hierbij zijn de deelnemers gevraagd om vanuit vier verschillende perspectieven (brillen) te kijken naar dreigingen. Deze aanpak is gekozen om een brede blik te faciliteren en de multidimensionaliteit van dreigingen te benadrukken en zo een aanzet te bieden voor discussie. Figuur 10 toont een voorbeeld van een risico van een AI-toepassing met mogelijke impact op de nationale veiligheid, vanuit de vier perspectieven: techniek, actoren, toepassingen en functies/doeleinden (van het AI-systeem zelf).

[29] Dit is in lijn met de risicobeoordelingsmethodiek die het ANV hanteert in de context van de Nationale Veiligheidsstrategie (NVS).
[30] NCTV, *Nationale Veiligheid Strategie 2019*, 9 (voetnoot 3).



Figuur 10: Voorbeelduitwerking van mogelijke risico's en dreigingen van AI-toepassingen

Hieronder wordt kort een aantal mogelijke dreigingen en risico's beschreven, zoals benoemd tijdens de expertsessie[31]:

- Mensen worden opgepakt op basis van de voorspelling dat ze een misdaad zullen begaan, waarbij inherent niet kan worden vastgesteld of diegene daadwerkelijk de misdaad zou zijn begaan als hij of zij niet zou zijn opgepakt.
- Gebruik van autonome, gewapende drones voor een militaire aanval op vitale infrastructuur.
- Apparatuur (bijvoorbeeld in kritieke overheidsprocessen) moedwillig laten oververhitten door een hack van besturingssystemen met behulp van AI.
- Verspreiding van desinformatie op het Internet middels geautomatiseerde processen.
- "*Weapons of math destruction*": het idee dat niet-transparante algoritmes volledig autonoom beslissingen maken met betrekking tot bijvoorbeeld kredietwaardigheid.
- "Kastenstelsel": de toepassingen van AI op big data kunnen ervoor zorgen dat mensen in bepaalde hokjes worden geplaatst op het gebied van bijvoorbeeld zorgpremies, opleidingen en hypotheken. Dit kan ertoe leiden dat een systeem van positieve terugkoppeling ontstaat waarbij men niet meer aan het profiel kan ontsnappen dat op basis van big data over hem of haar is gecreëerd.
- Controleverlies: technologieën met een dusdanig zelflerend vermogen dat mensen ze niet meer kunnen controleren, wat tot gevolg zou kunnen hebben dat autonome wapens de strijd met mensen of met elkaar aangaan.

[31] Tijdens de sessie zijn diverse risico's en dreigingen besproken maar is nog niet gekeken of deze in alle gevallen ook daadwerkelijk een aantasting van de nationale veiligheid kunnen veroorzaken.

Hierbij werd het voorbeeld van de *flash crash* genoemd, waarbij geautomatiseerde processen op elkaar reageerden en er binnen een paar seconden miljarden verdampten. Een militair equivalent hiervan zou betekenen dat autonome wapens op elkaar reageren en zo voor escalatie zorgen, op het gebied van het verstoren van communicatie of het daadwerkelijk aanrichten van fysieke schade.

- *Selling fear*: communicatie en informatieverstrekking over de impact van AI en de technologie zelf blijft achter, waardoor er in de samenleving doembeelden ontstaan van dreigingen die niet bestaan, dan wel niet relevant zijn. Kwaadwillende partijen kunnen hierop inspelen.
- Over-vertrouwen in AI: er kan te snel worden uitgegaan van de juiste werking van AI-systemen, maar er wordt niet beseft dat er vaak nog veel werk nodig is om een systeem echt goed te laten werken.
- AI-gedreven malware om *hacks* te plegen op bijvoorbeeld vitale infrastructuur.
- Afhankelijkheid van buitenlandse marktpartijen voor (gepatenteerde) sleuteltechnieken.
- Wetgeving loopt achter: capaciteit voor het ontwerpen en inregelen van wetgeving loopt achter op de snelle ontwikkeling van AI-technologie.
- 'Achterdeurtjes': AI-systemen die ingekocht zijn vanuit een marktpartij die niet transparant is over de werking van het systeem zorgen voor een strategische afhankelijkheid en de mogelijkheid voor de externe partij om bijvoorbeeld data af te tappen (denk ook aan de discussie rondom 5G en Huawei).
- *Adverserial* AI: het manipuleren van een AI-systeem of de data die het systeem moet voeden om zo andere uitkomsten te krijgen.

4.3 Risico-/dreigingscategorieën binnen het thema AI in de context van nationale veiligheid

Op basis van de literatuurverkenning en de output van de eerste expertsessie is door het projectteam een analyse gemaakt van archetypische risico's en dreigingen van AI-toepassingen. In de literatuur worden veel verschillende dreigingen en risico's van AI-toepassingen onderkend (van zeer uiteenlopende aard en orde) en in dit project is getracht om deze concrete mogelijke dreigingen en risico's op een hoger niveau te clusteren. In Tabel 1 staan 13 categorieën van risico's en dreigingen (in de tweede kolom), verspreid over 5 overkoepelende thema's (in de eerste kolom). De derde kolom geeft een korte uitleg wat de categorie inhoudt.

Thema	Categorie	Uitleg
Storing (*malfunction*)	1. Verlies van controle	Een AI-gedreven systeem raakt onbestuurbaar, of reageert onvoorspelbaar.
	2. Systeeminterferentie	AI-systemen reageren op elkaar, wat leidt tot onverwachte effecten.
	3. Verkeerde besluiten	AI-systemen maken verkeerde (niet-bedoelde) keuzes.
	4. Systeembeïnvloeding	Een actor verandert het gedrag van een AI-systeem door de

Gebruik voor kwaadwillende doeleinden		omgeving te veranderen, of door data te manipuleren.
	5. Emulatie	Een actor gebruikt AI-systemen om menselijk gedrag te emuleren, bijvoorbeeld in spraak, beeld of tekst.
	6. Overname van controle	Een actor neemt de controle van (netwerken van) AI-systemen over.
Toegenomen capaciteit van kwaadwillenden	7. Versterkte capaciteiten	AI-systemen versterken de bestaande capaciteiten van (kwaadwillende) actoren.
	8. Nieuwe capaciteiten	AI-systemen geven (kwaadwillende) actoren nieuwe capaciteiten.
Nieuwe verhoudingen	9. Diepe afhankelijkheid	Platformbedrijven en databedrijven beheersen de AI-technologiemarkt, en creëren monopolie posities.
	10. Lock-in	Gebruikers kunnen niet afstappen van hun AI-systemen omdat ze afhankelijk zijn van architecturen, data, standaarden.
	11. Beschikbaarheid en toegankelijkheid	AI-modellen en -algoritmes zijn openlijk beschikbaar, inclusief publieke datasets die gebruikt kunnen worden om AI-systemen te trainen.
	12. Nederlandse positie in de wereld	Nederland neemt posities in ten aanzien van ontwikkelingen op het gebied van AI. De ethische, wettelijke en marktposities die Nederland inneemt kunnen zorgen voor conflicten met andere staten, of de Nederlandse (economische) positie op het internationale toneel verzwakken.
Maatschappelijk	13. Publieke opinie	De Nederlandse samenleving kan zich keren tegen het gebruik van AI-gedreven systemen, of juist een overmatig vertrouwen hebben in de systemen, dan wel in regulering en controle.

Tabel 1: 13 risico-/dreigingscategorieën per thema

Het is belangrijk om te onderkennen dat de komst van AI-technologie het (nationale) veiligheidsdomein zal gaan veranderen, dat tot op zekere hoogte ook al heeft gedaan. Het gaat hierbij niet alleen om nieuwe risico's en dreigingen, maar ook reeds onderkende en bekende dreigingen kunnen door de komst van een nieuwe technologie worden beïnvloed. De dreiging van autonome wapensystemen is op zichzelf bijvoorbeeld geen nieuwe dreiging, aangezien militaire fysieke dreigingen al jaren worden aangemerkt als dreigingen voor de nationale veiligheid. Dus, de komst van AI kan in potentie bestaande dreigingen beïnvloeden. AI-technologie kan een versterkend (of wellicht juist dempend) effect hebben op 'bekende' risico's en dreigingen. De notie dat AI, als technologie, een totaal nieuwe impact heeft op veiligheid is daarom wellicht te kort door de bocht.

4.4 Mogelijke risico's en dreigingen binnen het thema AI in de context van nationale veiligheid

De benoemde risico-/dreigingscategorieën zijn bedoeld om mogelijke risico's en dreigingen te ordenen. Binnen elke categorie zijn (oneindig) vele concrete dreigingen en risico's denkbaar. De structuur van deze 13 categorieën (gebaseerd op de verkenning en eerste expertsessie) is in de tweede expertsessie gebruikt als kapstok om verder in te gaan op specifieke risico's en dreigingen die van belang zijn voor het beleidsterrein van nationale veiligheid.

De beleidsmedewerkers die deel hebben genomen aan de tweede expertsessie hebben, aan de hand van de categorieën en op basis van de concept map, nagedacht over concrete risico's en dreigingen. Hen werd gevraagd concrete huidige en toekomstige (tot tien jaar vooruit) dreigingen en risico's voor de nationale veiligheid te benoemen op basis van hun eigen ervaring en expertise. In potentie hebben veel diverse dreigingen en risico's impact op de nationale veiligheid, maar in deze sessie is gevraagd naar die dreigingen en risico's die bij de betrokken departementen (in de context van nationale veiligheid) op het netvlies staan. Hieronder volgt dezelfde tabel als hierboven, maar dan met concrete voorbeelden van dreigingen of risico's zoals die door de beleidsmedewerkers zijn benoemd. Voor elke categorie is één voorbeeld uitgewerkt, behalve voor categorie 12: 'Nederlandse positie in de wereld'. Dit werd in de expertsessie niet als een daadwerkelijk risico of dreiging gezien.

Vervolgens is aan de deelnemers gevraagd om een indicatie te geven welk van de categorieën voor de deelnemers het meest in het oog springen op basis van de concrete risico's en dreigingen die binnen die categorie zijn geïdentificeerd. De volgorde in Tabel 2 geeft de gradatie aan in belang. Hoewel deze gradatie slechts een representatie is van wat de deelnemersgroep ten tijde van de sessie relevant vond, biedt dit wel een interessante weergave van onderwerpen waar het meest op aangehaakt wordt. Dit kan een indicatie zijn voor onderwerpen die in een concrete impactanalyse kunnen worden uitgewerkt.

Categorie	Uitleg	Voorbeeld
9. Diepe afhankelijkheid	Platformbedrijven en databedrijven beheersen de AI-technologiemarkt, en creëren monopolie posities.	Buitenlandse technologiebedrijven (mogelijk verbonden aan staten met een ander

		waardenstelsel) maken misbruik van ongebreidelde toegang tot data.
4. Systeem-beïnvloeding	Een actor verandert het gedrag van een AI-systeem door de omgeving te veranderen, of door data te manipuleren.	Grootschalige fraude door middel van de beïnvloeding van geautomatiseerde financiële systemen.
13. Publieke opinie	De Nederlandse samenleving kan zich keren tegen het gebruik van AI-gedreven systemen, of juist een overmatig vertrouwen hebben in de systemen, dan wel in regulering en controle.	De Nederlandse bevolking ervaart AI als ingewikkeld en ongrijpbaar door gebrekkige informatievoorziening/ educatie over het fenomeen (vanuit het Rijk en eventuele andere instanties). Hierdoor kan de bevolking van alles wijs gemaakt worden door zelfbenoemde experts met kwade bedoelingen.
5. Emulatie	Een actor gebruikt AI-systemen om menselijk gedrag te emuleren, bijvoorbeeld in spraak, beeld of in tekst.	Het gebruik van *deepfakes* om publieke opinie te beïnvloeden door bijvoorbeeld buitenlandse actoren.
6. Overname van controle	Een actor neemt de controle van (netwerken van) AI-systemen over.	Een statelijke actor neemt de controle over van het elektriciteitsnet (of andere vitale infrastructuur).
3. Verkeerde besluiten	AI-systemen nemen verkeerde (niet-bedoelde) keuzes.	Bestaande biases in de samenleving (data) zoals racisme worden ingebouwd in een risicoprofileringssysteem.
7. Versterkte capaciteiten	AI-systemen versterken de bestaande capaciteiten van (kwaadwillende) actoren.	De verspreiding van desinformatie door middel van automatisering (trollen en bots) gaat sneller en efficiënter.
2. Systeem-interferentie	AI-systemen reageren op elkaar, wat leidt tot onverwachte effecten.	Controlesystemen voor vitale infrastructuur

		zetten elkaar onbedoeld uit.
10. Lock-in	Gebruikers kunnen niet afstappen van hun AI-systemen omdat ze afhankelijk zijn van architecturen, data, standaarden.	Afhankelijkheid van de verborgen agenda van de (buitenlandse) leverancier van het AI-systeem.
1. Verlies van controle	Een AI-gedreven systeem raakt onbestuurbaar, of reageert onvoorspelbaar.	Een AI-gedreven adviessysteem van de overheid produceert onbegrijpelijke resultaten die directe invloed hebben op besluitvorming (bij bijvoorbeeld de politie).
8. Nieuwe capaciteiten	AI-systemen geven (kwaadwillende) actoren nieuwe capaciteiten.	Export van AI-technologie met betrekking tot surveillance en gedrags- en persoonsherkenning op grote schaal.
11. Beschikbaarheid en toegankelijkheid	AI-modellen en -algoritmes zijn openlijk beschikbaar, inclusief publieke datasets die gebruikt kunnen worden om AI-systemen te trainen.	Toename van aanvallen door '*low budget*' criminelen en terroristen die mogelijk gemaakt worden door AI-systemen: 'c*ybercrime as a service*'.
12. Nederlandse positie in de wereld	De ethische, wettelijke en marktposities die NLD inneemt kunnen zorgen voor conflicten met andere staten, of de Nederlandse positie verzwakken.	*Dit werd in de expertsessie niet als een daadwerkelijk risico, dan wel dreiging gezien.*

Tabel 2: Concrete risico's en dreigingen binnen de 13 risicocategorieën

De benoemde dreigingen zeggen nog niets over de daadwerkelijke impact op nationale veiligheid. Hiervoor is een gestructureerde impactanalyse nodig middels bijvoorbeeld de ANV-methodiek (beoordeling van de impact op de nationale veiligheidsbelangen op basis van concrete, afgebakende scenario's). De voorbeelden die in het project geïdentificeerd zijn kunnen worden gezien als beknopte dreigingsnarratieven, waarbij er een context is gegeven aan het risico of de dreiging. In deze narratieven komen verschillende elementen van de concept map terug, zoals sectoren, actoren, toepassingen en randvoorwaarden. Hiermee bieden ze een goed startpunt voor het uitwerken van een scenario ten behoeve van een impactanalyse. Hier wordt verder op ingegaan in hoofdstuk 6.

5 Vervolgonderzoeksvragen

Het project geeft een interessant inzicht in de breedte van het onderwerp AI. De concept map, zoals weergegeven in deze notitie, geeft de resultaten weer van een eerste breedtescan rondom het fenomeen Kunstmatige Intelligentie. Alhoewel niet uitputtend, inspireert de concept map al wel tot vervolgvragen. Hieronder volgt een aantal.

Verkenning actorcomplexen rondom AI-technologie. Zoals besproken in paragraaf 3.2.5 over actoren, zijn er vele actoren betrokken bij de ontwikkeling en toepassing van AI-technologie. Het is waardevol om deze 'complexen' in kaart te brengen en een beeld op te bouwen van het krachtenveld. Hiermee ontstaan mogelijkheden om de effecten van interventies en incentives te analyseren. Hiervoor zouden systeem-dynamische modelleringmethodes zoals MARVEL[32] ingezet kunnen worden.

Verdiepende studie naar gewenst versus ongewenst gebruik van AI. Misbruik en ongewenst gebruik van AI zijn relatieve begrippen. Bedrijven zetten AI in voor versterking van hun eigen continuïteit. Dit kan bij andere actoren (zoals overheden) overkomen als ongewenst gebruik (bijvoorbeeld monopolisering, uitsluiten van gebruikersgroepen), en leiden tot lastige vraagstukken over data-eigendom, privacy en ethisch gebruik van AI-toepassingen. Bovendien kunnen goede bedoelingen uiteindelijk in de keten ongewenste effecten hebben. In een vervolgonderzoek zou bestudeerd kunnen worden hoe gewenst en ongewenst gebruik van AI-technologie zich manifesteert, en wat de implicaties daarvan zijn voor de overheid.

Manifestaties van AI. Het identificeren van concrete manifestaties van AI-technologie in de nabije toekomst. De media stelt dat AI grote verandering teweeg gaat brengen. Maar wat betekent dat nou concreet? In welke gebieden/sectoren gaan we grote veranderingen zien, en wat is de scope daarvan? Op de korte termijn gaan er geen significante veranderingen plaatsvinden. We gaan vooral veel nieuwe producten zien, en nieuwe diensten die daarop gebaseerd zijn. Maar als we de tijdscope langer vooruit zetten, dan wordt het vooruitzicht lastiger. Wat betekent het nou concreet als we zeggen dat AI de wereld gaan veranderen? Gaan we nieuwe ecosystemen zien? Gaan we nieuwe verhoudingen zien tussen publiek en overheid? We hebben meer inzicht nodig hoe dat er nou uitziet. Wat voor manifestaties van AI gaan we zien op middellange (10 jaar) tot lange termijn (20 jaar), en waar zitten mogelijkheden om die manifestaties te beïnvloeden?

De waarde van informatie. AI gaat de waarde van informatie beïnvloeden – zowel negatief als positief. Desinformatiecampagnes zijn nu aan de orde van de dag, en gaan versneld worden door AI-technologie. Aan de andere kant gaat AI ons

[32] Erik J.A. van Zijderveld, "MARVEL – principles of a method for semi-qualitative system behaviour and policy analysis," *TNO Defence, Security and Safety* (augustus 2007), https://www.tno.nl/media/9516/def_alg_paper_marvel_sds_2007.pdf; Guido A. Veldhuis, Peter van Scheepstal, Etiënne Rouwette, Tom Logtens, "Collaborative problem structuring using MARVEL," *EURO J Decis Process* 3 (2015).

ook helpen om informatie te vinden die waar is, en relevant voor onze doeleinden. Dit houdt in dat onze verstandhouding met informatie gaat veranderen. We mogen niet zomaar meer informatie vertrouwen, en er moeten nieuwe manieren ontstaat om met informatie om te gaan. Wat voor handvaten hebben we (als overheid) om waarheid te waarborgen? Wat voor handvatten heeft de samenleving om desinformatie aan te pakken?

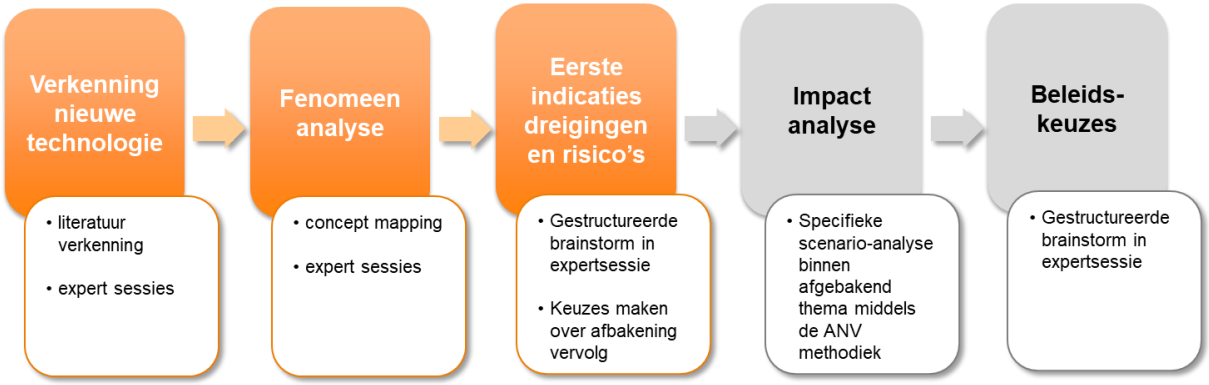
AI warfare: AI vs. AI. Veel ongewenste manifestaties van AI zullen gecounterd worden door AI-toepassingen. We krijgen mogelijk te maken met een situatie die op '*AI warfare*' gaat lijken: AI-systemen die door AI-systemen aangevallen worden. De snelheid en adaptief vermogen van AI-toepassingen kan alleen gebalanceerd worden door gelijkwaardige (AI-) toepassingen. We kunnen dit gaan zien op allerlei fronten. *Cyberwarfare*, desinformatie, Intelligence, verstoring en terrorisme zijn allemaal domeinen waarbinnen 'goede' AI-systemen de 'slechte' AI-systemen zullen gaan tegenkomen. Wat betekent dit voor controle (*meaningful human control*) en verantwoordelijkheid? Als de mens de respons overlaat aan AI-systemen, blijf je dan verantwoordelijk? Denk bijvoorbeeld eens aan een grootschalige digitale aanval op een ziekenhuis. Zou in zo'n geval een 'counter-AI'-systeem het ziekenhuis plat mogen leggen, of andere vitale systemen afschakelen? Met teveel restricties en wetgeving limiteer je de potentie van 'counter-AI-systemen', maar zonder restricties gaan er dingen mis. Hoe gaan we daarmee om, om Nederland veilig te houden?

Impactanalyse. Op basis van de concept map 'verkenning Kunstmatige Intelligentie in de context van nationale veiligheid' en de eerste aanzet tot het identificeren van mogelijke risico's en dreigingen van AI-toepassingen kan een keuze gemaakt worden voor een afgebakend scenario op basis waarvan een impactanalyse kan worden gedaan. De ANV methodiek, waarbij er een beoordeling wordt gemaakt van de impact op nationale veiligheidsbelangen, is hier geschikt voor, mits er een concreet, afgebakend scenario wordt gebruikt.

6 Methodiek – duiding nieuwe technologie

De in dit project gehanteerde aanpak zoals staat beschreven in hoofdstuk 2 biedt aanknopingspunten voor de duiding voor andere nieuwe technologieën, waarbij de nadruk voornamelijk ligt op het onderkennen van mogelijke dreigingen en risico's op de nationale veiligheid die voortvloeien uit de toepassing van nieuwe technologie. De aanpak zoals gebruikt in dit project zal in dit hoofdstuk worden veralgemeniseerd tot een methodiek die gebruikt kan worden om nieuwe technologieën of technologische ontwikkelingen te kunnen duiden in de context van nationale veiligheid.

Om een (nieuwe) technologie te kunnen duiden en op een niveau aan te komen dat er beleidskeuzes gemaakt kunnen worden, moeten er verschillende duidings- en analysestappen uitgevoerd worden. We beschrijven het proces in vijf elementaire stappen (zie Figuur 11) en lichten toe hoe deze aangepakt kunnen worden. De oranje blokken in onderstaand schema geven aan dat deze stappen tevens zijn uitgevoerd in dit project om AI in de context van nationale veiligheid te duiden.



Figuur 11: Processtappen duiding nieuwe technologie

1. Verkenning nieuwe technologie. In deze fase wordt een eerste brede verkenning van de technologie in kwestie uitgevoerd. Het doel van deze stap is om een initieel (en niet volledig) beeld te krijgen van het technologiegebied. Hierbij wordt gebruik gemaakt van literatuur en expertraadpleging, waarbij gestreefd wordt een zo integraal mogelijk beeld van het technologieveld op te bouwen. Het gaat hierbij niet zozeer om het streven naar inhoudelijke volledigheid (diepgang), maar om een goede structuur te krijgen die de relevante aspecten omvat (breedte). Daarom wordt niet alleen gekeken naar de technische aspecten van de technologie, maar ook bijvoorbeeld naar het gebruik van de technologie in producten, gerelateerde maatschappelijk ontwikkelingen, randvoorwaardelijke zaken voor het gebruik van de technologie en relevante actoren. De technologie wordt 'afgepeld' tot de voor de onderzoeksvraag relevante elementen, waarbij wordt gezocht naar een hogere ordening van deze elementen (die uiteindelijk als concepten op de concept map komen).

2. Fenomeenanalyse. Vanuit de verkenning wordt een fenomeenanalyse gedaan. Het doel van een fenomeenanalyse is om een gestructureerd beeld op te

bouwen van zaken die betrekking hebben op de technologie in kwestie. Deze elementen worden gedestilleerd uit de verkenning (stap 1) en in deze stap verder geordend en met elkaar in verband gebracht. Zo kunnen in deze stap clusters van concepten gemaakt worden of, bijvoorbeeld, vaak voorkomende relaties tussen technologievariaties en producten gelegd worden. Concept mapping als methode is in deze stap waardevol, maar andere technieken voor het vastleggen van een ordening van een technologie kunnen ook bruikbaar zijn.

3. Indicaties van risico's en dreigingen. De fenomeenanalyse in stap 2 leidt tot een overzicht van fenomenen die met de technologie in kwestie gemoeid zijn, al dan niet vastgelegd in een concept map. In deze fase van het project kunnen eerste indicaties van risico's en dreigingen afgeleid worden door verschillende concepten uit de ordening (concept map) te combineren en daarmee een context te creëren waarbinnen een risico geduid kan worden. De combinatie van concepten (uit de voorgaande stap) geeft context voor daadwerkelijke impactanalyse (in de volgende stap). Expertsessies met een diverse expertgroep zijn een goed middel om een breed palet aan mogelijke dreigingen en risico's te onderkennen.

4. Impactanalyse. Vanuit de mogelijke risico- en dreigingsbeschrijvingen kan er een analyse uitgevoerd worden naar de potentiële impact van specifieke risico's of dreigingen op de nationale veiligheid. De methodiek zoals die binnen het Analistennetwerk Nationale Veiligheid wordt gehanteerd[33] is hier uitermate geschikt voor, mits er een goede afbakening is gemaakt. De ANV methodiek (en impactanalyse in het algemeen) leent zich voornamelijk voor onderwerpen die gespecificeerd worden (bijvoorbeeld in een scenario, zoals binnen het ANV gebeurt). Doel van de impactanalyse is om een rangschikking te maken van risico's en dreigingen binnen het afgebakende thema op basis van impact op de nationale veiligheid en waarschijnlijkheid van optreden.

5. Beleidskeuzes. Op basis van de resultaten van de impactanalyse, en de onderliggende duiding van het technologieveld, kunnen er afgewogen beleidskeuzes op afgebakende onderwerpen vanuit het technologieveld gemaakt worden. Deze stap ligt inherent bij beleidsmedewerkers.

[33] ANV, *Leidraad risicobeoordeling. Geïntegreerde risicoanalyse Nationale Veiligheid*.

7 Conclusie

Gedurende de loop van het project is gebleken dat het zinnig en relevant is om de duiding van nieuwe technologieën op een gestructureerde manier aan te pakken. In de complexe, veranderende wereld anno nu speelt technologie een grote rol, maar hoe een nieuwe technologie impact gaat hebben op een samenleving en op veiligheid is vaak ongrijpbaar. In dit project is een aanzet gegeven tot een methodiek om een dergelijke duiding gedegen aan te pakken, zonder dat er jarenlange intensieve studies moeten worden uitgevoerd.

Met dit project is voor AI als nieuwe technologie een basis gelegd voor een nadere impactanalyse en beleidsvorming. Op basis van een literatuurverkenning is een fenomeenanalyse gedaan, waarvan de resultaten zijn geborgd in een concept map. Deze concept map is aangevuld, verrijkt en gevalideerd in twee expertsessies, waar zowel technisch-inhoudelijke experts als beleidsmedewerkers in dit vakgebied aan hebben deelgenomen. De concept map AI faciliteert een open blik om potentiële risico's en dreigingen te identificeren. De combinatie van verschillende elementen uit de map belichten de multidimensionale eigenschappen van mogelijke risico's en dreigingen voor de nationale veiligheid.

Op basis hiervan heeft dit project, naast de concept map, geleid tot een categorisering van mogelijke risico's en dreigingen met betrekking tot AI. Deze categorisering maakt het mogelijk om concrete risico's en dreigingen te identificeren, zonder dat een uitputtend overzicht van alle mogelijke risico's en dreigingen hoeft worden gemaakt. De verschijningsvormen van risico's en dreigingen zijn sterk verbonden aan toepassingsdomeinen en de context waarbinnen dit plaatsvindt. De risico- en dreigingscategorieën geven richting aan toekomstige analyses en maken het bovendien mogelijk oog te houden voor nieuwe specifieke manifestaties van deze typen risico's of dreigingen. De voorbeelden van risico's en dreigingen die tijdens het project geïdentificeerd zijn én de rangschikking van de categorieën die door de betrokken beleidsmakers is gemaakt, biedt een startpunt voor een diepgaandere impactanalyse die middels een afgebakend scenario kan worden uitgevoerd in een later traject. Wanneer deze impactanalyse is uitgevoerd, kunnen er gefundeerde beleidskeuzes worden gemaakt.

Bovendien is in dit project gebleken dat de gehanteerde methodiek om nieuwe technologie te duiden en concept mapping als tool aanzet geven tot verdere mogelijke onderzoeksvragen. Deze mogelijke onderzoeksvragen komen voort uit de discussies die inherent worden gevoerd tijdens de fenomeenanalyse en het concept mappen. In hoofdstuk 5 zijn aldus suggesties gegeven voor vervolgonderzoeksvragen, zowel op het gebied van diepergaande analyses over AI als op het gebied van nationale veiligheid.

Datum
27 januari 2020

Onze referentie
TNO 2020 M10094

Bijlage 1 – Overzicht bronmateriaal

Regio	Types	Actoren	Beschrijving	Relevante documenten	URL
Nederland	**Beleid**	Ministerie van Justitie en Veiligheid		Concept technologiescan JenV	
				Nationale Veiligheid Strategie 2019	
		Ministerie van Economische Zaken en Klimaat		Strategisch actieplan voor AI	
		Ministerie van Buitenlandse Zaken			
		Ministerie van Defensie		Defensie Industrie Strategie (samen met EZK)	https://www.defensie.nl/downloads/beleidsnota-s/2018/11/15/defensie-industrie-strategie
		Wetenschappelijke Raad voor het Regeringsbeleid		WRR: Internationaal AI-beleid. Domme data, slimme computers en wijze mensen	https://www.wrr.nl/publicaties/working-papers/2019/06/12/internationaal-ai-beleid
	Bedrijfs-leven en innovatie	ICAI: Innovation Center for AI	Nederlands netwerk van partners uit de academische wereld, de industrie en de overheid (onder meer Universiteit van Amsterdam, Nationale Politie, ING, TU Delft, Vrije Universiteit Amsterdam, Radboud Universiteit Nijmegen, Ahold, Qualcomm)	Lancering	https://www.uva.nl/content/nieuws/persberichten/2018/04/lancering-nationaal-innovation-center-for-ai.html
		FME Platform AI	Industriële ondernemersorganisatie voor de technologische industrie (onder meer: NXP, Siemens, IBM, ABB, DAF, Tata Steel, KPN, Philips, Thales, ASML)	Roadmap FME Platform Artificial Intelligence	https://www.fme.nl/system/files/publicaties/Roadmap%20FME%20Platform.pdf https://www.vno-ncw.nl/sites/default/files/position_paper_algoritmen_die_werken_voor_iedereen.pdf
		AINED	Samenwerking tussen het TopTeam ICT, VNO-NCW, ICAI,	AI voor Nederland : vergroten, versnellen en verbinden	https://www.vno-ncw.nl/sites/default/files/aivnl_20181106_0.pdf

			NWO en TNO, BCG, DenkWerk		
		Smart Industry	In november 2014 opgericht door het ministerie van Economische Zaken (TNO, KVK, KMU, FME, RVO), en voert sinds begin 2018 de Implementatieagenda 2018-2021 uit.		
		ECP Platform voor de Informatie-samenleving		Artificial Intelligence Impact Assessment	https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf
	Denktanks & analyse	Stichting Toekomst der Techniek		Duikboten zwemmen niet	
		PAX for Peace		Don't be evil - A survey of the tech sector's stance on lethal autonomous weapons	https://www.paxforpeace.nl/media/files/pax-report-killer-robots-dont-be-evil.pdf
		HCSS - Hague Center for Strategic Studies		Macro implications of micro transformations: an assessment of AI's impact on contemporary geopolitics	https://hcss.nl/report/macro-implications-micro-transformations-assessment-ais-impact-contemporary-geopolitics
		Analistennetwerk Nationale Veiligheid		Horizonscan Nationale Veiligheid 2018	https://www.clingendael.org/nl/publicatie/horizonscan-nationale-veiligheid-2018
		Rathenau Instituut			
		DenkWerk		Artificial Intelligence in Nederland. Zelf aan het stuur	https://denkwerk.online/media/1029/artitificial_intelligence_in_nederland_juli_2018.pdf
	Wetenschap	Universiteit Utrecht (CKI, TKI)		Cognitieve AI	
		Radbouw Universiteit Nijmegen		Sociale AI / Deep learning	
		VU Amsterdam - Artificial Intelligence		Machine learning / Robotica	

		UvA - Humane AI		Human AI	
		TU Delft - DAIRE Delft Artificial Intelligence Research and Education		Machine learning	
Europa	**Beleid**	EU		The European AI Alliance	https://ec.europa.eu/digital-single-market/en/european-ai-alliance
		EU		Coordinated Action Plan AI	https://ec.europa.eu/digital-single-market/en/european-ai-alliance
		EU		European Artificial Intelligence (AI) leadership, the path for an integrated vision	https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX:52018DC0795
		European Data Protection Supervisor		Declaration on Ethics and Data Protection in Artificial Intelligence	https://edps.europa.eu/sites/edp/files/publication/icdppc-40th_ai-declaration_adopted_en_0.pdf
		EC Joint Research Center		Artificial Intelligence: A European Perspective	http://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf
		High Level Expert Group on AI		High Level Expert Group on AI	https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
		European Commission, DG Connect, Robotics and AI unit		The European Artificial Intelligence landscape	https://ec.europa.eu/digital-single-market/en/news/european-artificial-intelligence-landscape
		Frankrijk		France AI Strategy - 'AI for Humanity'	https://www.aiforhumanity.fr/en/
		Duitsland		Künstliche Intelligenz (KI) ist ein Schlüssel zur Welt von morgen.	https://www.ki-strategie-deutschland.de/home.html
	Bedrijfs-leven en innovatie	Applied AI (DE)		Diverse white papers	https://appliedai.de/wp-content/uploads/2019/03/AppliedAI_Elements-of-a-comprehensive-AI-Strategy.pdf

		Thales		Launch European AI Platform	https://www.thalesgroup.com/en/group/journalist/press-release/launch-first-european-artificial-intelligence-platform-coordinated
	Denktanks & analyse	European Union Institute for Security Studies		AI - What implications for EU security and defence?	https://www.iss.europa.eu/content/artificial-intelligence-%E2%80%93-what-implications-eu-security-and-defence
		European Council on Foreign Relations		Machine politics: Europe and the AI revolution	https://www.ecfr.eu/publications/summary/machine_politics_europe_and_the_ai_revolution
	Wetenschap	Cambridge University		The facets of AI: A framework to track the evolution of AI	https://www.ijcai.org/proceedings/2018/0718.pdf
		Andere universiteiten			
		German Research Center for Artificial Intelligence (DFKI)			
		Fraunhofer Geselschafft			
		CNRS, INRIA, France			
		CLAIRE: Confederation of Laboratories for Artifical Intelligence Research in Europe		Vision document	https://claire-ai.org/wp-content/uploads/2018/09/CLAIRE-Vision-Document-2-2.pdf
		European Association for Artificial Intelligence: Home			
Interna-tionaal	**Beleid**	Verenigde Staten		Artificial Intelligence for the American People	https://www.whitehouse.gov/ai/
		US department of homeland security		AI: Using standards to mitigate risks	https://www.dhs.gov/sites/default/files/publications/2018_AEP_Artificial_Intelligence.pdf
		China		China Institute for Science and Technology Policy at Tsinghua University - China AI	http://www.sppm.tsinghua.edu.cn/eWebEditor/UploadFile/China_AI_development_report_2018.pdf

				Development Report 2018	
		Russia		Defense One - Russia Racing to Complete National AI Strategy by June 15	https://www.defenseone.com/threats/2019/03/russia-racing-complete-national-ai-strategy-june-15/155563/
		Russia		Defense One - Here's How the Russian Military Is Organizing to Develop AI	https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/
		OECD		OECD principles on AI	https://www.oecd.org/going-digital/ai/principles/
		G20		G20 Ministerial Statement on Trade and Digital economy	https://www.mofa.go.jp/files/000486596.pdf
		International Telecommunications Union		ITU AI Repository	https://www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx
	Bedrijfs-leven en innovatie	Facebook AI			
		Google AI			
		Alphabet/DeepMind			
		Microsoft AI			
		Samsung AI Research Lab			
		Partnership on AI			
	Denktanks & analyse	Center for Strategic and International Studies		AI and National Security	https://www.csis.org/analysis/artificial-intelligence-and-national-security-importance-ai-ecosystem
		Harvard Kennedy School Belfer Center		AI and National Security	https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf
		Center for a New American Security		AI and International security AI - What every	https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security

				policymaker needs to know	https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know
		Carnegie Endowment for International Peace		What the Machine Learning Value Chain Means for Geopolitics	https://carnegieendowment.org/2019/08/05/what-machine-learning-value-chain-means-for-geopolitics-pub-79631
		Future of Life Institute	The AI Initiative is an initiative of The Future Society		
		Allen AI			
	Wetenschap	Association for the Advancement of Artificial Intelligence (AAAI)		AAAI: A 20 year community roadmap for AI research in the US	https://aaai.org/Conferences/AAAI-19/townhall-a-20-year-roadmap-for-ai-research/
		OpenAI			
		International Joint Conferences on Artificial Intelligence (IJCAI)			
		CSAIL @ MIT	MIT Computer Science & Artificial Intelligence Lab		

Bijlage 2 – Organisaties van geraadpleegde experts

TNO
Stichting Toekomstbeeld der Techniek
Universiteit van Amsterdam
Rathenau Instituut
Ministerie van Justitie en Veiligheid
Ministerie van Defensie
Ministerie van Buitenlandse Zaken
Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
Rijksinstituut voor Volksgezondheid en Milieu / Secretariaat ANV

Datum
27 januari 2020

Onze referentie
TNO 2020 M10094

Blad
46/46

Bijlage 3 – Concept map verkenning Kunstmatige Intelligentie in de context van nationale veiligheid