



Raad van de
Europese Unie

Brussel, 26 oktober 2022
(OR. en)

14095/22
ADD 1

LIMITE

CORLX 995
CFSP/PESC 1432
CONUN 254
CODUN 48
CONOP 112
COTER 256
COARM 217

VOORSTEL

van:	de hoge vertegenwoordiger van de Unie voor buitenlandse zaken en veiligheidsbeleid, ondertekend door de heer Stefano SANNINO, secretaris-generaal
ingekomen:	26 oktober 2022
aan:	mevrouw Thérèse BLANCHET, secretaris-generaal van de Raad van de Europese Unie
Betreft:	Voorstel van de hoge vertegenwoordiger van de Unie voor buitenlandse zaken en veiligheidsbeleid aan de Raad voor een besluit van de Raad ter ondersteuning van de uitvoering van het project "Bevordering van verantwoorde innovatie in artificiële intelligentie voor vrede en veiligheid"

Hierbij gaat voor de delegaties document HR(2022) 238.

Bijlage: HR(2022) 238

EUROPESE DIENST VOOR EXTERN OPTREDEN



PROJECTDOCUMENT

Bevordering van verantwoorde innovatie in AI voor vrede en veiligheid

HR(2022)238

1. ACHTERGROND

Door de recente vooruitgang op het gebied van artificiële intelligentie (AI) zijn er nieuwe mogelijkheden ontstaan om vrede en veiligheid te ondersteunen en te handhaven, bijvoorbeeld door technologische verbeteringen op gebieden als vroegtijdige waarschuwing bij conflicten, en wapenbeheersing en wapenuitvoercontrole. Anderzijds heeft deze vooruitgang middelen voortgebracht die spanningen, conflicten en onveiligheid tussen en binnen staten veroorzaken of verergeren. De risico's van bepaalde AI-toepassingen, zoals dodelijke autonome wapensystemen, zijn een groot punt van zorg voor de wapenbeheersingsgemeenschap. Een risico dat meer aandacht verdient en dat momenteel misschien onvoldoende wordt aangepakt in het kader van wapenbeheersing en diplomatie, is de omleiding of het misbruik van civiel AI-onderzoek en civiele AI-innovatie, zoals misbruik van generative adversarial networks (GAN's) om deepfakes voor desinformatiecampagnes te maken, door onverantwoordelijke actoren, waaronder kwaadwillige niet-overheidsactoren.

AI is een ontsluitende technologie met een groot potentieel voor algemeen gebruik. Civiele toepassingen die op basis van AI-onderzoek en -innovatie zijn ontwikkeld, kunnen (relatief gemakkelijk) door bepaalde actoren worden aangewend en herbestemd voor schadelijke of disruptieve doeleinden die de internationale vrede en veiligheid in gevaar kunnen brengen. Omleiding en misbruik van civiele technologie zijn geen nieuwe verschijnselen en zijn evenmin uniek voor AI. Zo zijn er onlangs op het verwante gebied van robotica recreatieve drones door Daesh/ISIS in Syrië ingezet als wapens. In het geval van AI is het probleem echter gecompliceerder: door de niet-tastbare, snel veranderende aard van AI-algoritmen en -gegevens, is de overdracht/proliferatie ervan moeilijk te controleren; door de leidende rol van de privésector in het ecosysteem van onderzoek, ontwikkeling en innovatie moeten door eigendomsrechten beschermde algoritmen worden beschermd, en de menselijke expertise en materiële middelen om AI-technologieën te herbestemmen, zijn wereldwijd beschikbaar. Daarnaast zijn degenen die actief zijn in de civiele AI-sector zich al te vaak nog niet bewust van de mogelijke gevolgen die omleiding en misbruik van hun werkzaamheden kunnen hebben voor

de internationale vrede en veiligheid of zijn ze terughoudend om deel te nemen aan de lopende besprekingen over AI-risico's in wapenbeheersings- en non-proliferatiekringen.

De civiele AI-gemeenschap moet nauwer worden betrokken bij het onderzoeken en beperken van de risico's voor de vrede en veiligheid die gepaard gaan met omleiding en misbruik van civiele AI-technologie door onverantwoordelijke actoren. Volgens het Internationaal Instituut voor Vredesonderzoek van Stockholm (SIPRI) en het Bureau van de Verenigde Naties voor ontwapeningszaken (Unoda) kan dit worden bereikt door verantwoorde innovatie in de wereldwijde civiele AI-gemeenschap sterker te bevorderen. Uit eerdere werkzaamheden van het SIPRI en het Unoda blijkt dat verantwoorde innovatie, als een op zichzelf staand mechanisme, de wereldwijde civiele AI-gemeenschap praktische instrumenten en methoden zou kunnen aanreiken om de risico's vast te stellen van omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie voor de vrede en veiligheid, en om deze te voorkomen en beperken. Het SIPRI en het Unoda hebben bij hun werkzaamheden ook methoden en verschillende lopende initiatieven inzake verantwoorde AI voor de civiele gemeenschap vastgesteld waarop kan worden voortgebouwd om de civiele AI-gemeenschap bewust te maken van problemen in verband met wapenbeheersing en non-proliferatie en op de hoogte te brengen van deskundigendebatten en standpunten van staten over verantwoorde ontwikkeling, verspreiding en gebruik van AI. Ook is vastgesteld welke lessen kunnen worden getrokken uit de werkzaamheden van de defensiesector op het gebied van verantwoordelijkheid¹. Uit deze eerdere werkzaamheden is duidelijk naar voren gekomen dat samenwerking met studenten wetenschap, technologie, engineering en wiskunde (Science, Technology, Engineering and Mathematics, STEM), die nog via het onderwijs bezig zijn met AI, van cruciaal belang is voor doeltreffende verantwoorde innovatie-inspanningen.

2. DOELSTELLINGEN

Deze projecten zijn gericht op een grotere betrokkenheid van de civiele AI-gemeenschap bij het verminderen van de risico's van omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie door onverantwoordelijke actoren voor de internationale vrede en veiligheid. Om dit doel te bereiken moeten de projecten in de eerste plaats een beter inzicht geven in hoe beslissingen op het gebied van de ontwikkeling en verspreiding van AI-onderzoek en -innovatie het risico op omleiding en misbruik kunnen beïnvloeden, en op hun beurt risico's of kansen voor vrede en veiligheid kunnen genereren, en moeten zij op de tweede plaats verantwoorde innovatieprocessen, -methoden en -instrumenten bevorderen die kunnen bijdragen tot de vreedzame toepassing van civiele innovaties en de verantwoorde verspreiding van AI-kennis. Daartoe zullen de projecten capaciteitsopbouw-, onderzoeks- en samenwerkingsactiviteiten ondersteunen die i) de capaciteit binnen de wereldwijde civiele AI-gemeenschap zullen

¹ Tot de methoden behoren bijvoorbeeld de aanbevolen praktijken van het Institute of Electrical and Electronics Engineers (IEEE) voor het beoordelen van het effect van autonome en intelligente systemen op het welzijn van de mens (IEEE Std 7010-2020), en de lijst van de deskundigengroep op hoog niveau inzake AI voor de beoordeling van betrouwbare artificiële intelligentie (Assessment List for Trustworthy Artificial Intelligence, ALTAI). De initiatieven omvatten: het wereldwijd initiatief van het IEEE inzake de ethiek van autonome en intelligente systemen, het partnerschap inzake AI en het mondiale partnerschap inzake AI.

vergroten om de vredes- en veiligheidsrisico's van omleiding en misbruik van civiele AI door onverantwoordelijke actoren aan te pakken via verantwoorde innovatieprocessen, en ii) het verband versterken tussen de risicobeperkende inspanningen op het gebied van verantwoorde AI in de civiele sfeer en de inspanningen die al worden gedaan in de ontwapenings-, wapen-beheersings- en non-proliferatiegemeenschap op intergouvernementeel niveau. Cruciaal is dat de projectactiviteiten niet bedoeld zijn om nieuwe normen, beginselen of regelgeving vast te stellen, of anderszins binnen te dringen in gebieden die onder de bevoegdheid van landen vallen. In plaats daarvan zijn zij erop gericht civiel verantwoorde innovatie-inspanningen te leveren met betrekking tot risico's voor de vrede en de veiligheid in verband met omleiding en misbruik van civiele AI door onverantwoordelijke actoren, en voorlichting te geven over bestaande relevante intergouvernementele inspanningen.

Om de civiele AI-gemeenschap daadwerkelijk te bereiken en te beïnvloeden, wordt in de projecten een driedelige aanpak gevolgd, met als doel:

- (a) samen te werken met onderwijzenden – samen met geselecteerde onderwijzenden en ontwikkelaars van academische curricula werken aan de ontwikkeling en bevordering van onderwijsmateriaal dat kan worden gebruikt om in de opleiding van toekomstige AI-professionals (bv. in cursussen over AI-ethiek en verantwoorde innovatie) de aandacht te vestigen op de vredes- en veiligheidsrisico's als gevolg van omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie door onverantwoordelijke actoren;
- (b) samen te werken met studenten – geselecteerde STEM-studenten uit de hele wereld leren hoe vredes- en veiligheidsrisico's als gevolg van omleiding en misbruik van civiele AI-ontwikkeling door onverantwoordelijke actoren kunnen worden vastgesteld, voorkomen of beperkt in het onderzoeks- en innovatieproces of via andere bestuursprocessen, en
- (c) samen te werken met de AI-sector – samenwerken met beroepsverenigingen en normalisatie-instellingen zoals het Institute of Electrical and Electronics Engineers (IEEE) om i) opleidingsmateriaal op maat onder technische professionals te verspreiden en hen via activiteiten te betrekken; ii) de inzet van AI voor vredes- en veiligheidsdoeleinden te ondersteunen, en iii) de dialoog en informatie-uitwisseling te vergemakkelijken tussen deskundigen uit de academische wereld, de privésector en de overheid over hoe het risico van omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie door onverantwoordelijke actoren kan worden beperkt.

Met die aanpak bereiken de projecten alle geledingen van de AI-gemeenschap, waaronder niet alleen de huidige professionals, maar ook de toekomstige generaties. Op die manier wordt ook een samenwerking tot stand gebracht tussen de academische wereld, het bedrijfsleven en andere deskundigen, en wordt er door de oprichting van netwerken die al deze actoren omvatten voor gezorgd dat toekomstige inspanningen niet verloren gaan.

Voor de projecten wordt ook gebruikgemaakt van de verenigende krachten en ervaringen van het SIPRI en het Unoda om de AI-gemeenschap wereldwijd, en niet alleen belanghebbenden van de EU, te beïnvloeden. Het SIPRI en het Unoda bevinden zich in een unieke positie om AI-actoren uit heel Afrika, Azië-Stille Oceaan, Europa en Noord- en Zuid-Amerika te bereiken en onderlinge samenwerking te vergemakkelijken. Beide entiteiten hebben ook ervaring op andere

gebieden van wetenschap en technologie, zoals dat van de biotechnologie, waar zich door tweërlei gebruik en proliferatie soortgelijke problemen voordoen. De projecten hebben ook baat bij specifieke omstandigheden in de Europese Unie, zoals a) het bestaan van geavanceerde processen met meerdere belanghebbenden op het gebied van verantwoorde AI; b) het hoge niveau van samenwerking en deskundigheid op het gebied van ontwapening, wapenbeheersing en non-proliferatie in de EU; c) de diverse banden die academische, onderzoeks- en privé-organisaties in de EU hebben met andere regio's, met name in het Mondiale Zuiden, waarmee ook samenwerking wordt beoogd, en d) de diversiteit van nationaliteiten van studenten, onderwijzenden en ingenieurs aan universiteiten, in onderzoeksinstellingen en in de privésector.

Inclusie zal een centrale overweging zijn bij de uitvoering van de activiteiten in het kader van de projecten. Om de AI-gemeenschap doeltreffend te ondersteunen, wordt in de projecten erkend dat de AI-gemeenschap uit een breed scala aan actoren bestaat, en met name dat

- (a) gender een zeer relevante factor is. De projecten zullen dan ook wat gender betreft in overeenstemming zijn met de systeembrede strategieën voor gendermainstreaming en -gelijkheid van de VN. Alle activiteiten in het kader van de projecten moeten openstaan voor vrouwen, en vrouwen zullen worden aangemoedigd eraan deel te nemen, en
- (b) doorheen alle activiteiten zal er worden gestreefd naar de inclusie van personen met een handicap en redelijke aanpassing aan hun behoeften. Dit houdt onder meer in dat belemmeringen voor de deelname van personen met een handicap worden weggenomen en dat er maatregelen worden genomen om met personen met een handicap samen te werken en beter rekening te houden met hun inhoudelijke standpunten en ervaringen.

3. PROJECTEN

De drie projecten die hieronder staan beschreven, moeten elkaar aanvullen en ondersteunen. Een aantal onderdelen zullen de gehele periode van 36 maanden lopen.

Project 1 - Productie van onderwijs- en capaciteitsopbouw materiaal voor de civiele AI-gemeenschap

3.1. Doel van het project

Project 1 richt zich op het verschaffen van kennis en middelen voor civiele AI-actoren ten behoeve van de evaluatie en vermindering van de risico's die omleiden en misbruik van civiel AI-onderzoek en civiele AI-innovatie door onverantwoordelijke actoren met zich mee kunnen brengen voor de internationale vrede en veiligheid. Het project heeft tot doel onderwijs- en capaciteitsopbouw materiaal te produceren dat AI-professionals uit alle regio's, van alle niveaus en uit alle sectoren (met inbegrip van op AI gerichte opleiders, ontwikkelaars van leerplannen, STEM-studenten en AI-ingenieurs en -onderzoekers in de academische wereld en de particuliere sector) de nodige informatie en instrumenten zal verschaffen om:

- (a) inzicht te krijgen in hoe civiel AI-onderzoek en civiele AI-innovatie kunnen worden omgeleid en misbruikt op manieren die risico's kunnen inhouden voor de internationale

vrede en veiligheid, en hoe beslissingen op het gebied van de ontwikkeling en verspreiding van onderzoek en innovatie het risico op omleiding en misbruik kunnen vergroten of verminderen;

- (b) inzicht te krijgen in de inspanningen die al worden gedaan op het gebied van ontwapening, wapenbeheersing en non-proliferatie om de risico's van omleiding en misbruik van civiel onderzoek en civiele innovatie te beperken; en
- (c) verantwoorde innovatie toe te passen op een manier die het risico op omleiding en misbruik bij de ontwikkeling en verspreiding van onderzoek en innovatie beperkt.

3.1.1. Projectbeschrijving

Dit project zal bestaan uit drie aparte sets onderwijs- en capaciteitsopbouw materiaal.

- (a) Handboek (1) – Het handboek bundelt basiskennis en middelen voor AI-actoren om in het onderzoeks- en innovatieproces de risico's van omleiding en misbruik van civiele AI-technologie door onverantwoordelijke actoren te evalueren en te beperken. Er wordt besproken waarom en hoe beslissingen rond de ontwikkeling en verspreiding van onderzoek en innovatie het risico op omleiding en misbruik kunnen beïnvloeden, en op hun beurt risico's of kansen voor vrede en veiligheid kunnen genereren. Het handboek bevat ook relevante verplichtingen op het gebied van internationaal recht en uitvoercontrole, naast veiligheids- en beveiligingsoverwegingen die momenteel onderwerp van gesprek zijn in militaire, ontwapenings-, wapenbeheersings- en non-proliferatiekringen; en voorbeeldprocessen en -instrumenten om verantwoorde innovatie toe te passen, zoals technologie-effectbeoordelingsmethoden en modellen voor risicobeoordeling.
- (b) Podcastserie (~ 10) – Deze podcasts zullen fungeren als een toegankelijk en aantrekkelijk medium voor AI-actoren om te leren waarom en op welke manier verantwoorde AI-innovatieprocessen de internationale vrede en veiligheid kunnen ondersteunen door beperking van de risico's van omleiding en misbruik door onverantwoordelijke actoren. In de serie zullen belangrijke thema's worden besproken (bijv. het patroon van omleiding en misbruik van onderzoek en innovatie op het gebied van tweërlei/algemeen gebruik; humanitaire, strategische en politieke uitdagingen in verband met het mogelijke misbruik van civiel AI-onderzoek en civiele AI-innovatie; uitdagingen waarmee ontwapenings-, wapenbeheersings- en non-proliferatiekringen worden geconfronteerd bij de inspanningen om risico's te beperken; verantwoorde innovatie door middel van risicobeoordeling; naleving van de uitvoercontrole; risicobeperking door ontwerp; verantwoord publiceren; klantkennis; en ervaring met theoretische simulatieoefeningen) waarbij de interviews die het projectteam zal voeren met vertegenwoordigers van relevante gemeenschappen als uitgangspunt zullen worden genomen.
- (c) Blogserie (9-10) – Het team zal een blogpostserie samenstellen om zo meer zichtbaarheid te geven aan inspanningen die pogen de grenzen tussen "verantwoorde AI" in de civiele sfeer en wapenbeheersing en non-proliferatie te overschrijden. De blogserie zal een platform bieden om inzichten, ideeën en oplossingen te verspreiden met betrekking tot de identificatie en aanpak van risico's in verband met de omleiding en het misbruik van

civiele AI in het onderzoeks- en innovatieproces. De blog zal de diversiteit aan denkbeelden en perspectieven in de AI-sector weergeven.

Dit materiaal zal openbaar worden verspreid via de websites van de uitvoerende actoren, hun aanwezigheid op sociale media en via rechtstreekse communicatie met relevante academische entiteiten, civiele beroepsverenigingen op het gebied van AI en andere relevante groepen.

3.1.2. Verwachte resultaten van het project

Dit project zal naar verwachting een nieuwe serie materialen tot stand brengen waarmee civiele AI-professionals bewust kunnen worden gemaakt van a) hoe civiel AI-onderzoek en civiele AI-innovatie kunnen worden omgeleid en misbruikt op manieren die risico's kunnen inhouden voor de internationale vrede en veiligheid, b) hoe dergelijke risico's worden aangepakt door de ontwapenings-, wapenbeheersings- en non-proliferatiegemeenschap, en c) hoe AI-professionals verder kunnen bijdragen aan de beperking van dergelijke risico's door middel van verantwoorde innovatieprocessen.

Er wordt verwacht dat de civiele AI-sector hierdoor meer betrokken zal worden bij het verminderen van de risico's die omleiding en misbruik van civiele AI met zich mee kunnen brengen voor de internationale vrede en veiligheid; dat de capaciteit van technische professionals om zich bezig te houden met relevante processen in de ontwapenings-, wapenbeheersings- en non-proliferatiegemeenschap wordt vergroot, en dat er meer nieuwe doelgroepen bij worden betrokken die traditioneel geen deel uitmaken van de inspanningen op het gebied van ontwapenings- en non-proliferatieonderwijs.

Het materiaal zal naar verwachting ook helpen bij de uitvoering van de andere projecten en als basis dienen voor de activiteiten op het gebied van onderwijs en capaciteitsopbouw in het kader van project 2 en voor dialoog- en samenwerkingsactiviteiten in het kader van project 3.

Verwacht wordt dat deze activiteiten op hun beurt hun weerslag zullen hebben op de productie en verfijning van het materiaal. Een dergelijke iteratieve aanpak zal naar verwachting bijdragen tot de aanpak van potentiële belemmeringen voor de bevordering, verspreiding en toepassing van de activiteiten binnen de AI-gemeenschap, met inbegrip van kwesties in verband met taal, inhoud, context en beschikbaarheid, die de impact ervan op mondiaal niveau, met name in het Mondiale Zuiden, zouden kunnen verminderen.

3.2. Project 2 – Onderwijs- en capaciteitsopbouwactiviteiten voor toekomstige AI-professionals

3.2.1. Doel van het project

Het doel van project 2 is om ervoor te zorgen dat het probleem van omleiding en misbruik van civiel AI-onderzoek door onverantwoordelijke actoren wordt opgenomen in het onderwijs aan toekomstige generaties AI-professionals. Op de lange termijn zal dit ervoor zorgen dat de studenten wetenschap, technologie, engineering en wiskunde (STEM) die de toekomst van AI vormgeven, zich bewust zullen zijn van de negatieve gevolgen die de omleiding en het misbruik van hun werk door onverantwoordelijke actoren kunnen hebben voor de internationale vrede en

veiligheid en dat zij beschikken over de basisinstrumenten die nodig zijn om een dergelijk risico in het onderzoeks- en innovatieproces te identificeren en te beperken.

Dit project zal bestaan uit een reeks educatieve capaciteitsopbouwworkshops met opleiders en studenten in samenwerking met geselecteerde internationale universiteiten en actoren uit het bedrijfsleven. Het project is daarmee gericht op de ontwikkeling van capaciteitsopbouwactiviteiten die opleiders en ontwikkelaars van academische curricula zouden kunnen gebruiken in de opleiding van toekomstige AI-professionals (bijvoorbeeld cursussen over AI-ethiek en verantwoorde innovatie) en besteedt aandacht aan overwegingen omtrent de risico's op omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie door onverantwoordelijke actoren waarna deze twee onderwerpen worden bekeken in de bredere vredes- en veiligheidscontext. Via deze workshops wil het project ook een netwerk van geïnteresseerde opleiders, leerplanontwikkelaars en studenten in kaart brengen die de verspreiding en promotie van het projectonderwijsmateriaal en de activiteiten voor capaciteitsopbouw in de AI-onderwijsgemeenschap en onder AI-professionals zouden ondersteunen. Deze netwerkcomponent heeft tot doel de duurzaamheid van de projecten na hun onmiddellijke looptijd te waarborgen en nauwere banden te smeden ter ondersteuning van civiele technische samenwerking met het oog op ambitieuzere doelstellingen op het gebied van vrede, veiligheid, ontwapening en wapenbeheersing.

3.2.2. Projectbeschrijving

In het kader van dit project zal er een reeks workshops over onderwijs en capaciteitsopbouw worden georganiseerd met opleiders en studenten van geselecteerde universiteiten van over de hele wereld. Deze workshops bestaan uit een mix van lezingen en interactieve activiteiten die opleiders en studenten de kans bieden om na te denken over de manier waarop civiel AI-onderzoek en civiele AI-innovatie kunnen worden omgeleid en misbruikt op manieren die risico's kunnen inhouden voor de internationale vrede en veiligheid en hoe dergelijke risico's kunnen worden geïdentificeerd, voorkomen of beperkt in het onderzoeks- en innovatieproces of via andere governanceprocessen. Deze activiteiten zullen voortbouwen op eerdere kleinschaligere proefprojecten van het VN-Bureau voor ontwapeningszaken (Unoda), waarbij werd geëxperimenteerd met methoden om samen te werken met STEM-studenten en ze er bewust van te maken dat het belangrijk is rekening te houden met de bredere impact van hun werk alsook deskundigheid van buiten hun werkgebieden aan te trekken. Concreet zouden deze activiteiten bestaan uit:

- (a) *regionale workshops voor capaciteitsopbouw voor opleiders en studenten (4)* – tijdens de regionale workshops zullen er activiteiten worden uitgevoerd en gepromoot die opleiders kunnen gebruiken om de capaciteit van STEM-studenten op het gebied van verantwoorde AI-innovatie te vergroten, met bijzondere aandacht voor de wijze waarop de risico's op omleiding en misbruik van civiele AI-technologie door onverantwoordelijke actoren kunnen worden geëvalueerd en beperkt. Elke workshop zal worden georganiseerd met een in de EU gevestigde universiteit en een prominente universiteit uit een andere mondiale regio, waarbij een gevarieerde groep in de EU gevestigde deelnemers altijd wordt verbonden met een gevarieerde groep van buiten de EU. De workshops zullen

Latijns-Amerika en het Caribisch gebied, Noord-Amerika, Afrika en Azië en de Stille Oceaan bestrijken. Op deze manier kunnen er studenten (op master- en doctoraatsniveau) deelnemen uit heel de wereld, ook van het Mondiale Zuiden. De workshop wordt hoofdzakelijk in het Engels gehouden, maar waar mogelijk krijgen de deelnemers de mogelijkheid om deel te nemen aan activiteiten op basis van alternatieve taalgroepen; en

(b) *internationale workshop over duurzame capaciteitsopbouw (1)* – de workshop zal voortbouwen op de lessen die zijn geleerd van de regionale workshops en zal de uitwisseling van informatie en ervaringen tussen opleiders en geselecteerde studenten van de bij het project betrokken universiteiten vergemakkelijken. Tijdens de workshop zal worden besproken hoe de activiteiten en instrumenten die in de loop van het project zijn uitgewerkt, kunnen worden verfijnd en verspreid buiten de groep deelnemende universiteiten. Er zal ook worden besproken hoe ervoor kan worden gezorgd dat studenten, zodra ze tot de beroepsbevolking zijn toegetreden, betrokken blijven bij verantwoorde AI die omleidings- en misbruikrisico's voor de internationale vrede en veiligheid tegengaat.

De netwerken en aanwezigheid van het SIPRI en het Unoda in Afrika, Azië-Stille Oceaan, Europa, Noord- en Zuid-Amerika zullen waar nodig worden gebruikt om aspecten van de activiteiten te faciliteren en te ondersteunen.

3.2.3. Verwachte resultaten van het project

Het project zal naar verwachting modellen voor capaciteitsopbouw- en samenwerkings-activiteiten creëren die opleiders en ontwikkelaars van academische curricula als voorbeeld zouden kunnen gebruiken om toekomstige AI-professionals bewust te maken van de problemen van omleiding en misbruik van civiele AI door onverantwoordelijke actoren. Ook kunnen deze modellen worden gebruikt om toekomstige AI-professionals te leren hoe zij deze problemen kunnen helpen beperken door middel van verantwoorde innovatieprocessen. Na voltooiing van de projectactiviteiten zullen de deelnemers (opleiders, maar ook STEM-studenten) naar verwachting in staat zijn verantwoorde innovatie-instrumenten, -methoden en -concepten te gebruiken en te bevorderen waarmee de risico's op omleiding en misbruik bij de ontwikkeling en verspreiding van civiel AI-onderzoek en civiele AI-innovatie kunnen worden geïdentificeerd en beperkt.

De projectactiviteiten zullen naar verwachting ook met zich meebrengen dat er een netwerk ontstaat van opleiders, leerplanontwikkelaars en studenten die niet alleen de projectactiviteiten binnen de AI-onderwijs- en beroepsgemeenschappen promoten (bijv. tijdens conferenties van de Computational Intelligence Society van het IEEE), maar ook beschikbaar zullen zijn om met technische capaciteit bij te dragen aan op staatsniveau geleide internationale governance-processen (bijv. het proces rond het Verdrag inzake het gebruik van bepaalde conventionele wapens in de context van opkomende technologieën op het gebied van dodelijke autonome wapensystemen).

De korte- en langetermijnwaarde van deze activiteiten zal aan de hand van enquêtes vóór en na de activiteit worden aangetoond.

3.3. Project 3 – Faciliteren van de duurzame ontwikkeling, verspreiding en impact van verantwoorde innovatie in AI op langere termijn voor vrede en veiligheid

3.3.1. Doel van het project

Project 3 heeft tot doel de duurzame ontwikkeling, verspreiding en impact van verantwoorde innovatie in AI op langere termijn te faciliteren als middel om de risico's die omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie met zich mee kunnen brengen voor de vrede en veiligheid, te beperken. Het is de bedoeling dit te doen door middel van rondetafel-gesprekken met de AI-sector, dialogen met meerdere belanghebbenden, het opstellen van een openbaar verslag en gerichte verspreidingsactiviteiten. Het project moet ervoor zorgen dat de gegenereerde werkzaamheden, met name de activiteiten op het gebied van onderwijs, capaciteits-opbouw en samenwerking, de AI-gemeenschap in brede zin bereiken en raken, op alle niveaus (van studenten tot ingenieurs en andere AI-professionals), en over geografische, sectorale en andere grenzen heen reiken. Om de mogelijkheid te vergroten om een breed en diepgaand effect te sorteren, is het van essentieel belang om op dit gebied samen te werken met beroeps-organisaties, zoals het IEEE, en multidimensionale samenwerkingen aan te gaan met de hele academische wereld, het bedrijfsleven en andere deskundigen. Door dergelijke inspanningen zullen geïnteresseerde vertegenwoordigers van verschillende AI-gemeenschappen hun verantwoordelijkheid voor het probleem op gaan pakken en hun eigen mening willen geven over de manier waarop risicobeperkende maatregelen kunnen worden uitgevoerd en duurzaam kunnen worden bevorderd binnen de gehele wereldwijde AI-gemeenschap. Voor de langetermijnwaarde van het project voor landen, intergouvernementele organisaties en anderen is het ook belangrijk dat AI-professionals kunnen leren van en samenwerken met regeringsdeskundigen die betrokken zijn bij risicobeperking in de context van ontwapening, wapenbeheersing en non-proliferatie. In het kader van duurzaamheid is het ook van cruciaal belang om ervoor te zorgen dat de inzichten die uit de samenwerkingsactiviteiten voortvloeien, op passende wijze worden geanalyseerd, geconsolideerd en verspreid.

3.3.2. Projectbeschrijving

Dit project bestaat uit de volgende hoofdonderdelen:

- (a) dialogen met meerdere belanghebbenden over "verantwoorde AI-innovatie voor vrede en veiligheid" (tot en met 9) – in deze reeks virtuele dialoogbijeenkomsten zullen deskundigen uit de academische wereld, de onderzoekswereld, de particuliere sector en de traditionele wapenbeheersing uit de EU en daarbuiten worden samengebracht om van gedachten te wisselen over technologische trends die kunnen leiden tot verspreidings-, omleidings- en misbruikrisico's met gevolgen voor de internationale vrede en veiligheid;
 - i.
 - ii. over hoe risicobeperking kan worden bewerkstelligd door middel van verantwoorde innovatieprocessen, -methoden en -middelen, en over mogelijkheden en uitdagingen op het gebied van dialoog en kennisdeling tussen

- belanghebbenden, inclusief belanghebbenden die actief zijn in andere sectoren zoals de biologische en chemische industrie; en
- iii. over de potentiële waarde, het doel en de vorm van een zelfvoorzienend netwerk van deskundigen en dialoogactiviteiten. De deskundigengroep komt verschillende keren per jaar bijeen en werkt aan de organisatie van twee openbare evenementen voor de bredere gemeenschap.

Van de negen virtuele vergaderingen staan er twee open voor publiek, teneinde een bredere raadpleging mogelijk te maken.

(b) rondetafelgesprekken met de particuliere sector (tot en met 6) – deze reeks virtuele rondetafelgesprekken zal een dialoog op gang brengen met actoren die werken met verantwoorde AI-innovatieprocessen in de particuliere sector (bijv. partnerschap inzake AI) over de manier waarop zij kunnen bijdragen tot het minimaliseren van de risico's van omleiding en misbruik van civiele AI-technologieën door onverantwoordelijke actoren. Daarnaast kunnen zij onderzoeken welke stimulansen binnen de ontwikkeling van de particuliere sector mogelijk zijn om dat te bewerkstelligen. De onderwerpen zijn onder meer:

- i. de relevantie van de internationale veiligheids- en ontwapeningscontext voor de particuliere sector;
- ii. de juridische kaders waarin de ontwikkeling, uitrol en exploitatie van AI een rol spelen overal ter wereld;
- iii. hoe kan worden voortgebouwd op risicobeoordelingsmechanismen en andere maatregelen die deel uitmaken of kunnen gaan uitmaken van verantwoorde innovatieprocessen en nalevingsprogramma's van bedrijven; en
- iv. lessen die kunnen worden geleerd van andere industrieën, processen en kaders met betrekking tot wapenbeheersing (bijv. biologische en chemische industrieën).

(c) verslag over de perspectieven van de AI-gemeenschap met betrekking tot wapenbeheersing en risicobeperking in AI, gericht op AI- en wapenbeheersings-gemeenschappen (I) – in dit verslag zullen de bevindingen en aanbevelingen van het project worden vastgelegd en geconsolideerd in één referentiedocument dat gericht is op zowel in civiele sfeer verantwoorde AI-gemeenschappen als de wapenbeheersings-gemeenschappen. In het verslag wordt besproken hoe internationale vredes- en veiligheidsrisico's in verband met omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie kunnen worden geïdentificeerd, geëvalueerd en aangepakt.

(d) verspreidingsevenementen gericht op overleg en samenwerking met AI-gemeenschappen en wapenbeheersingsgemeenschappen (nog te bepalen) – het team zal mogelijkheden zoeken om de werkzaamheden en de daaruit voortvloeiende bevindingen en aanbevelingen gedurende het hele project te communiceren. De vorm van de evenementen en de inhoud van de presentaties worden afgestemd op de behoeften van de doelgroepen. Het kan daarbij gaan om vergaderingen van de Groep Conop, de vergadering van de Europese AI-alliantie; de groep regeringsdeskundigen inzake opkomende technologieën op het gebied van dodelijke autonome wapensystemen; de

werkgroep van agentschappen inzake AI (IAWG-AI); het initiatief "AI for Good" van de ITU; de jaarlijkse innovatiedialoog van het UNIDIR; en het Institute of Electrical and Electronics Engineers. Het team zal ook trachten bilateraal overleg te plegen met relevante belanghebbenden van de overheid, uit de academische wereld en uit de particuliere sector.

3.3.3. Verwachte resultaten van het project

Verwacht wordt dat dit project de basis zal leggen voor de duurzame ontwikkeling, verspreiding en impact van de verantwoorde innovatie van AI-processen waarmee verspreidings-, omleidings- en misbruikrisico's en de gevolgen daarvan voor de vrede en veiligheid worden aangepakt, ook na het verstrijken van het besluit van de Raad.

De dialoog met meerdere belanghebbenden zal naar verwachting als voorbeeld dienen voor informatie-uitwisseling en samenwerking op het gebied van risicobeperking, niet alleen binnen de wereldwijde AI-gemeenschap, maar ook tussen de verantwoorde AI-gemeenschap in civiele sfeer en de ontwapenings-, wapenbeheersings- en non-proliferatiegemeenschappen. Een dergelijk voorbeeld zou gebruikt kunnen worden om beleidsmakers vertrouwd te maken met belangrijke technologische en wetenschappelijke ontwikkelingen die relevant zijn voor de verantwoorde innovatie van AI en om technische doelgroepen vertrouwd te maken met de omgeving waarin beleidsmakers zich momenteel bevinden. Het project zal naar verwachting de duurzame betrekkingen en samenwerking bevorderen tussen geïnteresseerde actoren zowel binnen dezelfde gemeenschap als tussen deze verschillende gemeenschappen. Dergelijke heterogene netwerkeffecten zullen naar verwachting leiden tot een sterkere ontwikkeling en grootschaligere promotie van verantwoorde AI-innovatie voor vrede en veiligheid, ook na de looptijd van het project.

De dialoog met de particuliere sector zal de betrokkenheid van de particuliere AI-sector bij de identificatie, preventie en beperking van de vredes- en veiligheidsrisico's als gevolg van de omleiding en het misbruik van civiel AI-onderzoek en civiele AI-innovatie, naar verwachting vergroten en verdiepen. Het project zal de belangrijkste betrokkenen bij de processen in de particuliere sector meer inzicht geven in en meer verantwoordelijkheid geven voor de problemen die moeten worden aangepakt. Daarnaast heeft het tot doel de (bredere) invoering en toepassing van verantwoorde innovatieprocessen, -methoden en -middelen in bestaande mechanismen en procedures voor risicobeheer van bedrijven te faciliteren.

De dialoog met verschillende belanghebbenden en de rondetafelconferentie met de particuliere sector zullen naar verwachting ook inzicht verschaffen in een aantal belangrijke kwesties, waaronder a) de wijze waarop verantwoorde innovatiemethoden en -middelen verder kunnen worden verfijnd en ingezet om risico's in verband met omleiding en misbruik van civiel AI-onderzoek en civiele AI-innovatie vast te stellen, te voorkomen en te beperken; b) hoe AI-onderzoek en -innovatie positief kunnen worden gebruikt om vredes- en veiligheidsdoelstellingen te ondersteunen (bijv. toepassingen voor vroegtijdige waarschuwing bij conflicten en humanitaire bijstand); en c) hoe er meer dialoog en informatie-uitwisseling kan worden gefaciliteerd tussen de inspanningen op het gebied van risicobeperking die zijn ondernomen door

de verantwoorde AI-gemeenschap in civiele sfeer (d.w.z. diverse door het IEEE geleide initiatieven) en de initiatieven die reeds op intergouvernementeel niveau lopen op het gebied van ontwapening, wapenbeheersing en non-proliferatie.

Het verslag en de verspreidingsactiviteiten zullen de inzichten die op basis van de projecten 1, 2 en 3 zijn verkregen, analyseren, consolideren en verspreiden, en aldus helpen om de bevindingen en aanbevelingen van de projectactiviteiten binnen de wereldwijde AI-gemeenschap en binnen de beleidsgemeenschap te promoten. Zij zullen naar verwachting ook bijdragen tot het waarborgen van de duurzaamheid van het effect na de looptijd van de projecten.

4. DUUR

De totale duur van de uitvoering van de projecten wordt op 36 maanden geraamd.