



Europeiska
unionens råd

Bryssel den 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

FÖRSLAG

från:	Unionens höga representant för utrikes frågor och säkerhetspolitik, undertecknat av Stefano SANNINO, generalsekreterare
inkom den:	26 oktober 2022
till:	Thérèse BLANCHET, generalsekreterare för Europeiska unionens råd
Ärende:	Förslag, framlagt för rådet av unionens höga representant för utrikes frågor och säkerhetspolitik, till rådets beslut till stöd för genomförande av projektet "Främjande av ansvarsfull innovation inom artificiell intelligens för fred och säkerhet"

För delegationerna bifogas dokument – HR(2022) 238.

Bilaga: HR(2022) 238

EUROPEISKA UTRIKESTJÄNSTEN



PROJEKTDOKUMENT

Främjande av ansvarsfull innovation inom artificiell intelligens för fred och säkerhet

HR(2022)238

1. BAKGRUND

Den senaste tidens framsteg inom artificiell intelligens (AI) har skapat nya möjligheter att stödja och upprätthålla fred och säkerhet, till exempel genom tekniska förbättringar på områden som tidig varning för konflikter samt verifiering av vapen- och exportkontroll. Samtidigt har dessa framsteg inneburit nya möjligheter att skapa eller förvärra spänningar, konflikter och osäkerhet mellan och inom stater. Riskerna med vissa tillämpningar av AI, såsom dödliga autonoma vapensystem, har visat sig vara ett stort problem inom vapenkontrollsamfundet. En risk som bör ägnas större uppmärksamhet och som man med nuvarande vapenkontrollinsatser och diplomatiska insatser kanske inte har tillräcklig förmåga att hantera är avledning och missbruk av civil AI-forskning och AI-innovation som oansvariga aktörer, bland annat illvilliga icke-statliga aktörer, gör sig skyldiga till. Exempel på detta är missbruk av generativa adversariella nätverk (GAN) för att producera s.k. deepfake för desinformationskampanjer.

AI är en möjliggörande teknik med stor allmän användningspotential. AI-forskning och AI-innovation som utvecklas för civila tillämpningar skulle (relativt enkelt) kunna nås och användas av vissa aktörer för skadliga eller omstörtande ändamål som kan få konsekvenser för den internationella freden och säkerheten. Avledning och missbruk av civil teknik är inga nya fenomen, och de är inte heller unika för AI. På det närliggande området robotteknik exemplifierades detta nyligen av Daish/IS användning av drönare för fritidsbruk som vapen i Syrien. När det gäller AI kompliceras problemet emellertid av flera faktorer: AI-algoritmers och AI-datas immateriella och snabbt föränderliga karaktär, som gör det svårt att kontrollera överföringen eller spridningen av dem, den privata sektorns ledande roll i ekosystemet för forskning, utveckling och innovation och det därav följande behovet av att skydda proprietära algoritmer samt den globala tillgången till mänsklig expertis och materiella resurser som är i stånd att använda AI-tekniken för andra ändamål. Samtidigt känner de som arbetar med AI inom den civila sektorn alltför ofta inte till de potentiella konsekvenser som avledning och missbruk av deras arbete kan få för den internationella freden och säkerheten eller också är de tveksamma till

att delta i de diskussioner om riskerna med AI som förs i vapenkontroll- och icke-spridningskretsar.

Man bör stödja ett ökat engagemang från det civila AI-samfundet för att förstå och minska de freds- och säkerhetsrisker som är förknippade med oansvariga aktörers avledning och missbruk av civil AI-teknik. Enligt Stockholms internationella fredsforskningsinstitut (Sipri) och FN:s kontor för nedrustningsfrågor (Unoda) skulle detta kunna göras genom att man i högre grad främjar ansvarsfull innovation i det globala civila AI-samfundet. Sipris och Unodas tidigare arbete har visat att ansvarsfull innovation, som en mekanism för självstyre, skulle kunna förse det globala civila AI-samfundet med praktiska verktyg och metoder för att dels identifiera de risker som avledning och missbruk av civil AI-forskning och AI-innovation kan utgöra för freden och säkerheten, dels bidra till att förebygga och minska dem. Sipri och Unoda har i sitt arbete också identifierat metoder och flera pågående civilt inriktade initiativ för ansvarsfull AI som man skulle kunna bygga vidare på för att öka det civila AI-samfundets medvetenhet om frågor som rör vapenkontroll och icke-spridning, expertdebatter och staters ståndpunkter om ansvarsfull utveckling, spridning och användning av AI samt lärdomar som kan dras av försvarssektorns ansvarsarbete.¹ Inom ramen för detta tidigare arbete har det tydligt fastställts att kontakter med studerande inom naturvetenskap, teknik, ingenjörsvetenskap och matematik (STEM), som fortfarande sysslar med AI inom ramen för sin utbildning, är av central betydelse för alla effektiva och ansvarsfulla innovationsinsatser.

2. MÅL

Syftet med dessa projekt är att stödja ett ökat engagemang från det civila AI-samfundet för att minska den risk som oansvariga aktörers avledning och missbruk av civil AI-forskning och AI-innovation kan utgöra för internationella fred och säkerhet. Avsikten är att göra detta genom att för det första skapa större förståelse för hur beslut om utveckling och spridning av AI-forskning och AI-innovation kan påverka riskerna för avledning och missbruk och i sin tur kan medföra risker eller möjligheter för fred och säkerhet och för det andra främja ansvarsfulla förfaranden, metoder och verktyg för innovation som kan bidra till att säkerställa fredlig tillämpning av civila innovationer och ansvarsfull spridning av AI-kunskap. I detta syfte stöder projekten kapacitetssuppleggning, forskning och engagemang som kommer att i) öka kapaciteten inom det globala civila AI-samfundet för att omfatta och hantera de freds- och säkerhetsrisker som uppstår till följd av oansvariga aktörers avledning och missbruk av civil AI genom ansvarsfulla innovationsförfaranden, och ii) stärka kopplingen mellan riskreducerande insatser inom ansvarsfull användning av AI på det civila området och insatser som redan pågår på mellanstatlig nivå inom nedrustnings-, vapenkontroll- och icke-spridningssamfundet. Det är av avgörande betydelse att de inte syftar till att fastställa nya normer, principer eller regler eller på annat sätt inverkar på områden som omfattas av staternas behörighet. Avsikten med projekten är i stället att

¹ Metoderna omfattar till exempel rekommenderad praxis från *Institute of Electrical and Electronics Engineers* (IEEE) för bedömning av autonoma och intelligenta systems inverkan på människors välbefinnande (IEEE Std 7010-2020) och AI-expertgruppens bedömningslista för tillförlitlig artificiell intelligens för självbedömning (ALTAI). Initiativen omfattar IEEE:s globala initiativ om etik för autonoma och intelligenta system, partnerskapet för artificiell intelligens och det globala partnerskapet för artificiell intelligens.

utveckla ansvarsfulla civila innovationsinsatser som ska omfatta freds- och säkerhetsrisker som uppstår till följd av oansvariga aktörers avledning och missbruk av civil AI, och tillhandahålla utbildning om befintliga relevanta mellanstatliga insatser.

För att faktiskt nå ut till och påverka det civila AI-samfundet tillämpas i projekten en tredelad strategi som har följande syften:

- (a) Samarbete med utbildare – samarbeta med utvalda utbildare och utvecklare av akademiska kursplaner om utveckling och främjande av utbildningsmaterial som kan användas för att integrera hänsyn till de freds- och säkerhetsrisker som följer av att oansvariga aktörer avleder och missbrukar civil AI-forskning och AI-innovation i utbildningen av framtida yrkesmässiga användare av AI-system (t.ex. i kurser om AI-etik och ansvarsfull innovation).
- (b) Samarbete med studerande – låta utvalda studerande inom naturvetenskap, teknik, ingenjörsvetenskap och matematik (STEM) från hela världen stifta bekantskap med hur de freds- och säkerhetsrisker som uppstår till följd av oansvariga aktörers avledning och missbruk av civil AI-utveckling kan identifieras, förebyggas eller minskas i forsknings- och innovationsprocessen eller genom andra styrningsprocesser.
- (c) Samarbete med AI-branschen – samarbeta med yrkesorganisationer och standardiseringsorgan som *Institute of Electrical and Electronics Engineers* (IEEE) för att i) sprida skraddarsydd utbildningsmaterial och engagemangsaktiviteter till tekniska specialister, ii) stödja positiv användning av AI för fred och säkerhet samt iii) underlätta dialog och informationsutbyte mellan experter från den akademiska världen, den privata sektorn och myndigheter om hur man kan minska risken för att oansvariga aktörer avleder och missbrukar civil AI-forskning och AI-innovation.

En sådan strategi gör att projekten kan nå ut till AI-samfundet på alla nivåer – inte bara till nuvarande yrkesutövare utan även framtida generationer. Den möjliggör också samarbete mellan akademiska, industriella och andra silor och stöder hållbarheten i framtida insatser genom att den upprättar nätverk som korsar dessa gränser.

Projekten syftar också till att utnyttja Sipris och Unodas samlande kraft och erfarenhet för att påverka AI-samfundet globalt, inte bara EU:s intressenter. Sipri och Unoda har unika förutsättningar att nå ut till och underlätta kontakter mellan AI-aktörer från hela Afrika, Asien och Stillahavsområdet, Europa samt Nord- och Sydamerika. Båda organen har också erfarenhet av att arbeta inom andra vetenskapliga och tekniska områden som står inför liknande utmaningar i form av dubbla användningsområden och spridning, däribland bioteknik. Projekten syftar också till att dra nytta av de förhållanden som råder inom Europeiska unionen, såsom a) förekomsten av avancerade flerpartsprocesser för ansvarsfull användning av AI, b) den höga graden av engagemang och sakkunskap när det gäller frågor som rör nedrustning, vapenkontroll och icke-spridning i EU, c) den mångfald av kopplingar som akademiska organisationer, forskningsorganisationer och organisationer inom den privata sektorn i EU har till andra regioner, särskilt i den globala södern, som också kommer att vara ett viktigt mål för deltagande, och d) den nationella mångfalden bland studenter, utbildare och ingenjörer inom universitet, forskningsinstitut och den privata sektorn.

Inkludering kommer att vara en central fråga vid genomförandet av projektverksamheten. För att kunna stödja AI-samfundet på ett effektivt sätt tar projekten hänsyn till att AI-samfundet består av en rad olika aktörer och särskilt till följande:

- (a) Kön är en mycket viktig faktor. Därför kommer jämställdhetsperspektivet att integreras i linje med FN-systemets strategier för jämställdhetsintegrering och jämvikt. Deltagande av kvinnor i all verksamhet inom projektet kommer att uppmuntras och krävas.
- (b) Inkludering av personer med funktionsnedsättning och skälig behovsanpassning kommer att ske under hela projektiden. Detta omfattar att man undanröjer hinder för att personer med funktionsnedsättning ska kunna delta och vidtar åtgärder för att samarbeta med dem och ta bättre hänsyn till deras synpunkter och erfarenheter i sak.

3. PROJEKT

De tre projekt som beskrivs nedan är avsedda att komplettera och stödja varandra, med delar som pågår under samtliga 36 månader.

Projekt 1 – Framställning av utbildnings- och kapacitetsuppbyggnadsmaterial för det civila AI-samfundet

3.1. Projektets syfte

Projekt 1 är inriktat på att tillhandahålla kunskap och medel för att civila AI-aktörer ska kunna utvärdera och minska de risker som oansvariga aktörers avledning och missbruk av civil AI-forskning och AI-innovation kan medföra för internationell fred och säkerhet. Syftet är att ta fram utbildnings- och kapacitetsuppbyggnadsmaterial för att förse yrkesmässiga användare av AI-system från alla regioner, nivåer och sektorer (inbegripet AI-inriktade utbildare, de som utvecklar kursplaner, studerande inom naturvetenskap, teknik, ingenjörsvetenskap och matematik samt AI-ingenjörer och AI-forskare inom den akademiska världen och den privata sektorn) med den information och de verktyg som krävs för att

- (a) förstå hur civil AI-forskning och AI-innovation skulle kunna avledas och missbrukas på sätt som kan medföra risker för internationell fred och säkerhet och hur beslut om utveckling och spridning av forskning och innovation kan öka eller minska risken för avledning och missbruk,
- (b) förstå de insatser som nedrustnings-, vapenkontroll- och icke-spridningssamfundet redan deltar i för att minska riskerna för avledning och missbruk av civil forskning och innovation, och
- (c) tillämpa ansvarsfull innovation på ett sätt som minskar risken för avledning och missbruk vid utveckling och spridning av forskning och innovation.

3.1.1. Projektbeskrivning

Tre separata uppsättningar utbildnings- och kapacitetsuppbyggnadsmaterial kommer att tas fram inom ramen för projektet.

- (a) Handbok (1) – I handboken kommer grundläggande kunskaper och metoder att sammanställas som gör det möjligt för AI-aktörer att i forsknings- och innovationsprocessen utvärdera och minska riskerna för oansvariga aktörers avledning och missbruk av civil AI-teknik. Man kommer att ta upp varför och hur beslut om utveckling och spridning av AI-forskning och AI-innovation kan påverka riskerna för avledning och missbruk och i sin tur kan medföra risker eller möjligheter för fred och säkerhet. Handboken kommer också att beskriva relevanta skyldigheter enligt internationell rätt och inom ramen för exportkontroll, liksom skydds- och säkerhetshänsyn som diskuteras inom militären samt inom nedrustnings-, vapenkontroll- och icke-spridningskretsar. Det kommer också att ges exempel på processer och verktyg för att tillämpa ansvarsfull innovation, till exempel metoder för teknisk konsekvensbedömning och mallar för riskbedömning.
- (b) Poddserie (~ 10) – Dessa poddsändningar kommer att utgöra ett tillgängligt och engagerande medium för AI-aktörer att lära sig varför och hur ansvarsfulla innovationsprocesser inom AI kan stödja internationell fred och säkerhet genom att minska riskerna med avledning och missbruk från oansvariga aktörers sida. Serien kommer att ta upp viktiga teman (t.ex. mönstret för avledning och missbruk av forskning och innovation med dubbel/allmän användning, humanitära, strategiska och politiska utmaningar i samband med potentiellt missbruk av civil AI-forskningsinnovation, utmaningar som nedrustnings-, vapenkontroll- och icke-spridningskretsar står inför i samband med riskreducerande insatser, hur ansvarsfull innovation kan praktiseras genom riskbedömning, efterlevnad av exportkontroll, inbyggd riskreducering, ansvarsfull publicering, kundkännedom, och erfarenhet av skrivbordsövningar) och kommer att struktureras kring intervjuer som projektteamet kommer att genomföra med företrädare för berörda grupper.
- (c) Bloggserie (9–10) – Teamet kommer att utveckla en kuraterad bloggserie med syftet att synliggöra insatser som försöker korsa gränserna mellan civilt fokuserade ”ansvarsfulla AI-grupper” och vapenkontroll- och icke-spridningsgrupper. Bloggserien kommer att utgöra en plattform för att sprida insikter, idéer och lösningar när det gäller identifiering och hantering av risker i samband med avledning och missbruk av civilt AI i forsknings- och innovationsprocessen. Bloggen kommer att försöka återge mångfalden av tankar och perspektiv inom AI-sektorn.

Materialet kommer att spridas till allmänheten via de genomförande parternas webbplatser, deras närvaro i sociala medier och genom direkt kommunikation med relevanta akademiska enheter, civila yrkessammanslutningar inom AI och andra lämpliga grupper.

3.1.2. Förväntade projektresultat

Detta projekt förväntas leda till en ny uppsättning material genom vilka civila AI-utövare kan bli medvetna om a) hur civil AI-forskning och AI-innovation skulle kunna avledas och missbrukas på sådana sätt som kan utgöra risker för internationell fred och säkerhet, b) hur dessa risker hanteras av nedrustnings-, vapenkontroll- och icke-spridningssamfundet och c) hur AI-utövare

ytterligare skulle kunna bidra till att minska dessa risker genom ansvarsfulla innovationsprocesser.

Detta väntas främja engagemanget från den civila AI-sektorn för att minska den risk som avledning och missbruk av civil AI kan utgöra för internationella fred och säkerhet, förbättra de tekniska aktörernas kapacitet att samverka i relevanta processer inom nedrustnings-, vapenkontrolls- och icke-spridningssamfundet samt stödja engagemanget hos nya målgrupper som traditionellt inte omfattas av utbildningsinsatserna för nedrustning och icke-spridning.

Materialet förväntas också stödja genomförandet av de andra projekten och kommer att ligga till grund för utbildnings- och kapacitetsuppbyggnadsverksamheten i projekt 2 samt för dialog- och engagemangsaktiviteterna i projekt 3. Denna verksamhet förväntas i sin tur bidra till produktionen och förfiningen av materialet. En sådan iterativ strategi förväntas bidra till att hantera potentiella hinder för främjande, spridning och användning av materialet inom AI-samfundet, inbegripet frågor som rör språk, innehåll, sammanhang och tillgänglighet, vilket skulle kunna förhindra materialets inverkan på global nivå, särskilt i den globala södern.

3.2. Projekt 2 – Utbildning och kapacitetsuppbyggnad för framtida AI-utövare

3.2.1. Projektets syfte

Syftet med projekt 2 är att bidra till att problemet med oansvariga aktörers avledning och missbruk av civil AI-forskning tas upp i utbildningen av framtida generationer av AI-utövare. På lång sikt kommer detta att säkerställa att studerande inom naturvetenskap, teknik, ingenjörsvetenskap och matematik (STEM) som formar AI:s framtid är medvetna om de negativa konsekvenser som oansvariga aktörers avledning och missbruk av deras arbete skulle kunna få för internationell fred och säkerhet och att de kommer att ha de grundläggande verktyg som krävs för att identifiera och minska en sådan risk i forsknings- och innovationsprocessen.

Inom detta projekt kommer det att genomföras en rad workshoppar om utbildning och kapacitetsuppbyggnad med utbildare och studerande i samarbete med utvalda internationella universitet och branschaktörer. Projektet syftar därmed till att utveckla kapacitetsuppbyggande verksamhet som utbildare och utvecklare av akademiska kursplaner skulle kunna inkludera i utbildningen av framtida AI-utövare (t.ex. kurser om AI-etik och ansvarsfull innovation) och överväganden om riskerna för att oansvariga aktörer avleder och missbrukar civil forskning och innovation inom AI och koppla dessa överväganden till den bredare freds- och säkerhetskontexten. Genom dessa workshoppar kommer man inom projektet också att försöka kartlägga ett nätverk av intresserade utbildare, kursplansutvecklare och studerande som skulle stödja spridningen och marknadsföringen av projektutbildningsmaterialet och kapacitetsuppbyggnadsverksamheten inom AI-utbildningssamfundet och grupper av AI-utövare. Denna nätverkskomponent syftar till att säkerställa projektens hållbarhet även efter det omedelbara genomförandet och till att göra det möjligt att bygga starkare band till stöd för ett civilt tekniskt engagemang för bredare mål för fred, säkerhet, nedrustning och vapenkontroll.

3.2.2. Projektbeskrivning

Inom detta projekt kommer det att genomföras en rad workshoppar om utbildning och kapacitetsuppbyggnad med utbildare och studerande från utvalda universitet världen över. Dessa skulle bestå av en blandning av föreläsningar och interaktiva aktiviteter som ger lärare och studerande möjlighet att reflektera över hur civil AI-forskning och AI-innovation skulle kunna avledas och missbrukas på sådana sätt som kan utgöra risker för internationell fred och säkerhet och hur dessa risker kan identifieras, förhindras eller minskas i forsknings- och innovationsprocessen eller genom andra styrningsprocesser. Dessa aktiviteter kommer att bygga på ett föregående, mer begränsat pilotarbete som utförts av Unoda, som experimenterade med metoder för att samarbeta med STEM-studerande och göra dem mer medvetna om vikten av att beakta de bredare effekterna av deras arbete och bygga upp kontakter med experter utanför det egna ämnesområdet. Konkret skulle dessa bestå av följande:

- (a) Regionala workshoppar för kapacitetsuppbyggnad för lärare och studerande (4) – De regionala workshopparna kommer att genomföra och främja verksamhet som utbildare kan använda för att bygga upp de STEM-studerandenas kapacitet när det gäller ansvarsfull AI-innovation, med särskilt fokus på hur man kan utvärdera och minska riskerna för att oansvariga aktörer avleder och missbrukar civil AI-teknik.
- (b) Varje workshop kommer att anordnas med ett EU-baserat universitet och ett högprofilerat universitet från en annan global region, vilket innebär att en rad olika EU-baserade deltagare alltid kopplas samman med olika deltagargrupper som är baserade utanför EU. Workshopparna kommer sedan att omfatta Latinamerika och Karibien, Nordamerika, Afrika, Asien och Stilla-havsområdet. Detta gör det möjligt för studerande (på master- och doktorandnivå) från hela världen, även från den globala södern, att delta. Workshopen skulle i första hand genomföras på engelska, men när så är möjligt skulle deltagarna ges möjlighet att delta i aktiviteter som bygger på alternativa språkgrupper.
- (c) Internationell workshop om hållbar kapacitetsuppbyggnad (1) – Workshopen kommer att bygga på lärdomar från de regionala workshopparna och underlätta utbytet av information och erfarenheter mellan lärare och utvalda studerande från de universitet som deltar i projektet. Vid workshopen skulle man diskutera hur man kan finslipa de aktiviteter och verktyg som utarbetats under projektets gång och sprida dem utanför gruppen av deltagande universitet. Man kommer också att diskutera hur man kan stödja de studerandes engagemang i ansvarsfullt AI som tar itu med avlednings- och missbruksrisker för internationell fred och säkerhet när de väl har kommit in på arbetsmarknaden.

Sipris och Unodas nätverk och närvaro i Afrika, Asien, Stilla-havsområdet, Europa, Nordamerika och Sydamerika kommer vid behov att användas för att underlätta och stödja olika aspekter av verksamheten.

3.2.3. Förväntade projektresultat

Projektet förväntas skapa modeller för kapacitetsuppbyggnad och engagemang som utbildare och utvecklare av akademiska kursplaner skulle kunna reproducera för att uppmärksamma framtida AI-utövare på problemen med oansvariga aktörers avledning och missbruk av civilt AI och hur

de kan bidra till att minska dessa problem genom ansvarsfulla innovationsprocesser. När projektverksamheten har slutförts förväntas deltagarna (lärare, men även studerande inom naturvetenskap, teknik, ingenjörsvetenskap och matematik) kunna använda och främja ansvarsfulla verktyg, metoder och koncept för innovation för att identifiera och minska riskerna för avledning och missbruk i samband med utveckling och spridning av civil AI-forskning och AI-innovation.

Projektverksamheten förväntas också skapa ett nätverk av utbildare, kursplansutvecklare och studerande som inte bara skulle främja projektverksamheten inom AI-utbildning och AI-yrkesgrupper (t.ex. under konferenser i IEEE Computational Intelligence Society), utan också vara tillgängligt för att bidra med teknisk kapacitet till statligt ledda internationella styrningsprocesser (t.ex. processen kring konventionen om vissa konventionella vapen när det gäller ny teknik på området dödliga autonoma vapensystem).

Dessa aktiviteters värde på kort och lång sikt kommer att påvisas genom undersökningar före och efter aktiviteterna.

3.3. Projekt 3 – Främja långsiktigt hållbar utveckling, spridning och inverkan av ansvarsfull innovation inom AI för fred och säkerhet

3.3.1. Projektets syfte

Syftet med projekt 3 är att underlätta långsiktigt hållbar utveckling, spridning och inverkan av ansvarsfull innovation inom AI som ett sätt att minska de risker som avledning och missbruk av civil AI-forskning och AI-innovation kan utgöra för fred och säkerhet. Målet är att göra detta genom rundabordssamtal med AI-branschen, flerpartsdialoger, utarbetande av en offentlig rapport och riktad spridningsverksamhet. Projektet syftar till att säkerställa att det arbete som genereras, särskilt utbildnings-, kapacitetsuppbyggnads- och engagemangsaktiviteter, når och påverkar AI-samfundet i stort, på alla nivåer (från studerande till ingenjörer och andra AI-utövare) och över geografiska, sektoriella och andra gränser. För att öka möjligheterna att få en bred och djupgående inverkan är det viktigt att samarbeta med yrkesorganisationer på detta område, såsom IEEE, och att få till stånd flerdimensionella möten mellan den akademiska världen, industrin och andra silor. Dessa insatser kommer att ge berörda företrädare från olika AI-samfund möjlighet till egenansvar för problemet och att lämna sina egna synpunkter på hur riskreduceringsinsatser kan genomföras och främjas på ett hållbart sätt inom och över hela det globala AI-samfundet. Det är också viktigt för projektets långsiktiga värde för stater, mellanstatliga organisationer och andra att AI-utövare kan lära av och samarbeta med regeringsexperten som arbetar med riskreducering i samband med nedrustning, vapenkontroll och icke-spridning. Det är också avgörande för hållbarheten att säkerställa att de insikter som genereras genom kontaktverksamheten analyseras, konsolideras och sprids på lämpligt sätt.

3.3.2. Projektbeskrivning

Detta projekt består av följande centrala delar:

- (a) *Flerpartsdialoger om ansvarsfull AI-innovation för fred och säkerhet (upp till 9) –*
Denna serie virtuella dialogmöten skulle samla experter från den akademiska världen,

forskningsvärlden och den privata sektorn och experter på traditionell vapenkontroll från EU och övriga världen för att diskutera

- i. teknologiska trender som kan skapa risker vad avser spridning, avledning och missbruk med konsekvenser för internationell fred och säkerhet,
- ii. hur man kan bidra till riskreducering genom ansvarsfulla processer, metoder och medel för innovation samt möjligheter och utmaningar för dialog och kunskapsutbyte mellan berörda parter, inbegripet de som är verksamma inom andra sektorer, t.ex. den biologiska och kemiska industrin, och
- iii. det potentiella värdet av, syftet med och formatet för ett självbärande nätverk av experter och dialogaktiviteter. Expertgruppen kommer att sammanträda flera gånger per år och arbeta för att anordna två offentliga evenemang för samhället i stort.

Av de nio virtuella mötena är två avsedda att vara öppna för allmänheten för att underlätta ett bredare samråd.

- (b) Rundabordssamtal med aktörer inom den privata sektorn (upp till 6) – Denna serie virtuella rundabordssamtal kommer att inleda en dialog med aktörer som arbetar med ansvarsfulla innovationsprocesser inom AI inom den privata sektorn (t.ex. partnerskap för AI) om hur de kan bidra till att minimera riskerna för oansvariga aktörers avledning och missbruk av civil AI-teknik samt utforska möjliga incitament inom ramen för den privata sektorns utveckling för att göra detta. Ämnena omfattar följande:
- i. Betydelsen av internationell säkerhet och nedrustning för den privata sektorn.
 - ii. Den rättsliga miljö eller de rättsliga miljöer där AI redan utvecklas, införs och används runt om i världen.
 - iii. Hur man kan bygga vidare på riskbedömningsmekanismer och andra åtgärder som ingår i, eller skulle kunna integreras i, ansvarsfulla innovationsprocesser och program för företags regelefterlevnad.
 - iv. Erfarenheter från andra branscher, processer och ramar som rör vapenkontroll (t.ex. den biologiska och kemiska industrin).
- (c) Rapport om AI-samfundets syn på vapenkontroll och riskreducering inom AI, med inriktning på AI-samfundet och vapenkontrollsamfundet (1) – Framställningen av denna rapport kommer att fånga upp och befästa projektets resultat och rekommendationer i ett enda referensdokument som riktar sig till både det ansvarsfulla AI-samfundet inom den civila sfären och vapenkontrollsamfundet. I rapporten kommer man att ta upp hur internationella freds- och säkerhetsrisker i samband med avledning och missbruk av civil AI-forskning och AI-innovation kan identifieras, utvärderas och hanteras.
- (d) Spridningsevenemang som är inriktade på samråd och kontakter med AI-samfundet och vapenkontrollsamfundet (bekräftas senare) – Projektteamet kommer att söka möjligheter att informera om arbetet och dess resultat och rekommendationer under hela projektiden. Evenemangens format och presentationernas innehåll kommer att anpassas till målgruppernas behov. Det kan bland annat handla om möten i arbetsgruppen för icke-spridning och vapenexport, församlingen för Europeiska AI-alliansen, gruppen av

regeringsexperten på ny teknik på området dödliga autonoma vapensystem, FN:s arbetsgrupp om AI (IAWG-AI), ITU:s initiativ ”AI for Good”, Unidirs årliga innovationsdialog och *Institute of Electrical and Electronics Engineers*. Projektgruppen kommer också att sträva efter att samarbeta bilateralt med relevanta intressenter från myndigheter, den akademiska världen och den privata sektorn.

3.3.3. Förväntade projektresultat

Detta projekt förväntas lägga grunden för hållbar utveckling, spridning och inverkan av ansvarsfull innovation inom AI-processer som motverkar risker för spridning, avledning och missbruk och deras konsekvenser för fred och säkerhet även efter det att rådets beslut har löpt ut.

Flerpartsdialogen förväntas utgöra en modell för informationsutbyte och samarbete om riskreducering, inte bara inom det globala AI-samfundet utan även mellan det ansvarsfulla AI-samfundet inom den civila sfären och nedrustnings-, vapenkontroll- och icke-spridningssamfundet. En sådan modell skulle kunna användas för att göra beslutsfattare förtroga med viktiga tekniska och vetenskapliga framsteg som är relevanta för ansvarsfull AI-innovation och även för att göra tekniska målgrupper förtroga med den miljö som beslutsfattarna för närvarande verkar i. Projektet förväntas underlätta hållbara förbindelser och kontakter mellan berörda aktörer inom och mellan dessa olika samfund. Sådana heterogena nätverkseffekter förväntas göra det möjligt att ytterligare utveckla och i högre grad främja ansvarsfull AI-innovation för fred och säkerhet även efter projektets slut.

Dialogen med den privata sektorn förväntas möjliggöra ett större och djupare engagemang från den privata AI-sektorn i identifieringen, förebyggandet och minskningen av de freds- och säkerhetsrisker som härrör från avledning och missbruk av civil AI-forskning och AI-innovation. Projektet förväntas ge nyckelaktörer inom den privata sektorns processer större förståelse och egenansvar för de problem det är avsett att lösa. Dessutom syftar det till att underlätta ett (bredare) antagande och genomförande av ansvarsfulla processer, metoder och medel för innovation inom befintliga mekanismer och förfaranden för företags riskhantering.

Flerpartsdialogen och rundabordssamtalet med aktörer inom den privata sektorn förväntas också generera insikter i ett antal viktiga frågor, bland annat a) hur ansvarsfulla metoder och medel för innovation kan vidareutvecklas och användas för att identifiera, förebygga och minska riskerna med avledning och missbruk av civil AI-forskning och AI-innovation, b) hur AI-forskning och AI-innovation kan användas på ett positivt sätt för att stödja freds- och säkerhetsmål (t.ex. tillämpningar för tidig varning för konflikter och humanitärt bistånd) och c) hur man kan främja mer dialog och informationsutbyte mellan de riskreducerande insatser som görs inom det ansvarsfulla AI-samfundet inom den civila sfären (dvs. olika IEEE-ledda initiativ) och de som redan pågår på mellanstatlig nivå inom nedrustnings-, vapenkontroll- och icke-spridningssamfundet.

Rapporten och spridningsverksamheten kommer att analysera, konsolidera och sprida insikter som genererats genom projekten 1, 2 och 3 och därmed stödja främjandet av resultat och rekommendationer från projektverksamheten inom det globala AI-samfundet och inom det

politiska samfundet. De förväntas också bidra till att säkerställa hållbara effekter även efter projektens slut.

4. VARAKTIGHET

Genomförandet av projekten beräknas pågå under totalt 36 månader.