



Rådet for
Den Europæiske Union

Bruxelles, 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

FORSLAG

fra:	Stefano SANNINO, generalsekretær, på vegne af Unionens højtstående repræsentant for udenrigsanliggender og sikkerhedspolitik
modtaget:	26. oktober 2022
til:	Thérèse BLANCHET, generalsekretær for Rådet for Den Europæiske Union
Vedr.:	Forslag fra Unionens højtstående repræsentant for udenrigsanliggender og sikkerhedspolitik til Rådet med henblik på Rådets afgørelse til støtte for gennemførelse af et projekt med titlen "Fremme af ansvarlig innovation inden for kunstig intelligens for fred og sikkerhed"

Hermed følger til delegationerne dokument HR(2022) 238.

Bilag: HR(2022) 238

TJENESTEN FOR EU'S OPTRÆDEN UDADTIL



PROJEKTDOKUMENT

Fremme af ansvarlig innovation inden for kunstig intelligens for fred og sikkerhed

HR(2022) 238

1. BAGGRUND

Nylige fremskridt inden for kunstig intelligens (AI) har åbnet nye muligheder for at støtte og opretholde fred og sikkerhed, for eksempel gennem teknologiske forbedringer på områder såsom tidlig varsling af konflikter samt verifikation af våben- og eksportkontrol. På den anden side har disse fremskridt gjort det muligt at skabe eller forværre spændinger, konflikter og usikkerhed mellem og i stater. De risici, der er forbundet med visse anvendelser af AI, såsom dødbringende automatvåbensystemer, har vist sig at være et stort problem for våbenkontrolsamfundet. En risikovej, der fortjener større opmærksomhed, og som den nuværende våbenkontrolmæssige og diplomatiske indsats kan være utilstrækkelig til at reagere på, er uansvarlige aktørers, herunder ondsindede ikkestatslige aktørers, omdirigering og misbrug af civil AI-forskning og -innovation, f.eks. misbrug af generative adversarial networks (GAN) til at producere "deepfakes" til desinformationskampagner.

AI er en støtteteknologi med stort potentiale inden for generel anvendelse. AI-forskning og -innovation, der udvikles til civile anvendelser, kan (relativt let) tilgås og anvendes til andre formål af visse aktører med henblik på skadelige eller disruptive anvendelser, som kan have konsekvenser for den internationale fred og sikkerhed. Omdirigering og misbrug af civil teknologi er ikke nye fænomener og er heller ikke enestående for AI. På det relaterede område robotteknologi blev dette for nylig eksemplificeret ved Daesh/ISIS' anvendelse af rekreative droner som våben i Syrien. Men i forbindelse med AI kompliceres problemet af mange faktorer: AI-algoritmers og -datas immaterielle og hurtigt skiftende karakter, som gør det vanskeligt at kontrollere overførsel/spredning af dem, den private sektors ledende rolle i forsknings-, udviklings- og innovationsøkosystemet og det deraf følgende behov for at beskytte proprietære algoritmer samt den globale tilgængelighed af menneskelig ekspertise og materielle ressourcer, der er i stand til at ændre anvendelsen af AI-teknologier til andre formål. Samtidig er de, der arbejder inden for AI i den civile sektor, stadig for ofte ikke klar over de potentielle konsekvenser, som omdirigering og misbrug af deres arbejde kan have for den internationale fred og sikkerhed, eller tøver med at deltage i de aktuelle drøftelser om AI-risici i våbenkontrol- og ikkespredningskredse.

Der er behov for at støtte en større inddragelse af det civile AI-miljø i at forstå og begrænse de trusler mod fred og sikkerhed, der er forbundet med uansvarlige aktørers omdirigering og misbrug af civil AI-teknologi. Ifølge Det Internationale Fredsforskningsinstitut i Stockholm (SIPRI) og De Forenede Nationers Kontor for Nedrustningsanliggender (UNODA) kan dette opnås ved i højere grad at fremme ansvarlig innovation i det globale civile AI-miljø. SIPRI's og UNODA's tidligere arbejde har vist, at ansvarlig innovation som en selvstyringsmekanisme kan give det globale civile AI-miljø praktiske redskaber og metoder til at identificere samt bidrage til at forebygge og begrænse den trussel, som omdirigering og misbrug af civil AI-forskning og -innovation kan udgøre for fred og sikkerhed. SIPRI's og UNODA's arbejde har også identificeret metoder og flere igangværende civilt fokuserede initiativer vedrørende ansvarlig AI, der kan bygges på for at bevidstgøre det civile AI-miljø om våbenkontrol- og ikkespredningsspørgsmål, ekspertdebatter og statslige holdninger til ansvarlig udvikling, udbredelse og anvendelse af AI samt erfaringer, som skal indhøstes fra arbejdet med forsvarssektorens ansvar¹. Kritisk har dette tidligere arbejde klart vist, at inddragelse af studerende inden for naturvidenskab, teknologi, ingeniørvirksomhed og matematik (STEM), som stadig beskæftiger sig med AI inden for uddannelse, er af central betydning for enhver effektiv ansvarlig innovationsindsats.

2. MÅL

Disse projekter har til formål at støtte øget inddragelse af det civile AI-miljø i mindskelsen af de risici, som uansvarlige aktørers omdirigering og misbrug af civil AI-forskning og -innovation kan udgøre for international fred og sikkerhed. De sigter mod at gøre dette ved for det første at skabe større forståelse af, hvordan beslutninger om udvikling og udbredelse af AI-forskning og -innovation kan påvirke risiciene for omdirigering og misbrug og dermed skabe trusler mod eller muligheder for fred og sikkerhed, og for det andet ved at fremme ansvarlige innovationsprocesser, -metoder og -værktøjer, der kan bidrage til at sikre fredelig anvendelse af civile innovationer og ansvarlig udbredelse af AI-viden. Med henblik herpå støtter de kapacitetsopbygnings-, forsknings- og inddragelsesaktiviteter, der skal i) øge kapaciteten i det globale civile AI-miljø til at tage hensyn til og håndtere de trusler mod fred og sikkerhed, som uansvarlige aktørers omdirigering og misbrug af civil AI frembyder, gennem ansvarlige innovationsprocesser og ii) styrke forbindelsen mellem arbejdet med at mindske risici inden for ansvarlig AI på det civile område og det allerede igangværende arbejde inden for nedrustnings-, våbenkontrol- og ikkespredningsfællesskabet på mellemstatsligt plan. Det er afgørende, at de ikke har til hensigt at fastsætte nye standarder, principper eller bestemmelser eller på anden måde bevæge sig ind på områder, der henhører under staternes kompetence. I stedet har de til hensigt at udvikle arbejdet inden for civil ansvarlig innovation for at tage hensyn til de trusler mod fred og sikkerhed, som uansvarlige aktørers omdirigering og misbrug af civil AI frembyder, og sørge for uddannelse i eksisterende relevant mellemstatsligt arbejde.

For effektivt at nå ud til og påvirke det civile AI-miljø anvender projekterne en trestrengt tilgang, der søger at

¹ Metoderne omfatter for eksempel den anbefalede praksis fra Institute of Electrical and Electronics Engineers (IEEE) for vurdering af autonome og intelligente systemers indvirkning på menneskers trivsel (IEEE Std 7010-2020) og Ekspertgruppen på Højt Niveau vedrørende Kunstig Intelligens' vurderingsliste for pålidelig kunstig intelligens (ALTAI). Initiativerne omfatter IEEE's globale initiativ om etik i forbindelse med autonome og intelligente systemer, partnerskabet om AI og det globale partnerskab om AI.

- (a) samarbejde med undervisere – samarbejde med udvalgte undervisere og udarbejdere af akademiske læseplaner om udarbejdelse og fremme af undervisningsmateriale, som kan anvendes til at integrere overvejelser om de trusler mod fred og sikkerhed, der skyldes uansvarlige aktørers omdirigering og misbrug af civil AI-forskning og -innovation, i uddannelsen af fremtidige AI-aktører (f.eks. i kurser om AI-etik og ansvarlig AI-innovation)
- (b) samarbejde med studerende – introducere udvalgte studerende inden for naturvidenskab, teknologi, ingeniørvirksomhed og matematik (STEM) fra hele verden for, hvordan de trusler mod fred og sikkerhed, som uansvarlige aktørers omdirigering og misbrug af civil AI-udvikling frembyder, kan identificeres, forebygges eller begrænses i forsknings- og innovationsprocessen eller gennem andre styringsprocesser, og
- (c) samarbejde med AI-branchen – samarbejde med faglige sammenslutninger og standardiseringsorganer såsom Institute of Electrical and Electronics Engineers (IEEE) om at i) udbrede skræddersyede undervisningsmaterialer og inddragelsesaktiviteter til tekniske fagfolk, ii) støtte positive anvendelser af AI med henblik på fred og sikkerhed og iii) lette dialog og informationsudveksling mellem eksperter fra den akademiske verden, den private sektor og regeringen om, hvordan risikoen for uansvarlige aktørers omdirigering og misbrug af civil AI-forskning og -innovation kan begrænses.

En sådan tilgang gør det muligt for projekterne at nå ud til AI-miljøet på alle niveauer, herunder ikke blot nuværende aktører, men også fremtidige generationer. Den muliggør også inddragelse på tværs af akademiske, industrielle og andre siloer og støtter bæredygtigheden af fremtidige bestræbelser ved at oprette netværk, der krydser disse grænser.

Projekterne søger desuden at benytte sig af SIPRI's og UNODA's indkaldelsesbeføjelser og erfaringer til at påvirke AI-miljøet globalt og ikke kun EU-interessenter. SIPRI og UNODA har enestående forudsætninger for at nå ud til og lette engagement mellem AI-aktører fra Afrika, Asien og Stillehavsområdet, Europa og Nord- og Sydamerika. De to enheder har også erfaring med at arbejde inden for andre videnskabelige og teknologiske områder, der står over for lignende udfordringer vedrørende dobbelt anvendelse og spredning, herunder bioteknologi. Projekterne søger ydermere at drage fordel af forhold, der findes i Den Europæiske Union, såsom a) eksistensen af avancerede multiinteressentprocesser for ansvarlig AI, b) det høje niveau af engagement i og ekspertise inden for nedrustnings-, våbenkontrol- og ikke-spredningsspørgsmål i EU, c) mangfoldigheden af forbindelser, som akademiske organisationer, forskningsorganisationer og organisationer i den private sektor i EU har med andre regioner, navnlig i det globale Syd, som også vil være et vigtigt mål for inddragelse, og d) mangfoldigheden af nationaliteter blandt studerende, undervisere og ingeniører på universiteter og i forskningsinstitutioner og den private sektor.

Inklusion vil være et centralt hensyn i forbindelse med gennemførelsen af projektaktiviteterne. For effektivt at støtte AI-miljøet anerkender projekterne, at AI-miljøet består af en bred vifte af aktører, og navnlig, at

- (a) køn er en yderst relevant faktor. Derfor vil kønsaspektet blive integreret i overensstemmelse med FN-systemets strategier for integration af kønsaspektet og ligestilling. Deltagelse af kvinder i alle aktiviteter i projektet vil blive fremmet og krævet, og
- (b) inklusion af personer med handicap og rimelig imødekommelse af behov vil blive gennemført i hele forløbet. Dette vil omfatte overvindelse af hindringer for deltagelse af personer med

handicap samt sikring af, at der tages skridt til at søge kontakt med og lette tilkendegivelse af væsentlige synspunkter og erfaringer fra personer med handicap.

3. PROJEKTER

Det er hensigten, at de tre projekter, der er beskrevet nedenfor, skal være komplementære og gensidigt understøttende med elementer, som løber i alle de 36 måneder.

Projekt 1 – Udarbejdelse af uddannelses- og kapacitetsopbygningsmateriale til det civile AI-miljø

3.1. Projektformål

Projekt 1 fokuserer på at tilvejebringe viden og midler til civile AI-aktører med henblik på at evaluere og begrænse den trussel, som uansvarlige aktørers omdirigering og misbrug af civil AI-forskning og -innovation kan udgøre mod international fred og sikkerhed. Det har til formål at udarbejde uddannelses- og kapacitetsopbygningsmateriale, som vil give AI-aktører fra alle regioner, niveauer og sektorer (herunder AI-fokuserede undervisere, udarbejdere af læseplaner, STEM-studerende samt AI-ingeniører og -forskere i den akademiske verden og den private sektor) de oplysninger og redskaber, der er nødvendige for at

- (a) forstå, hvordan civil AI-forskning og -innovation kan omdirigeres og misbruges på måder, der kan udgøre en trussel mod international fred og sikkerhed, og hvordan beslutninger inden for udvikling og udbredelse af forskning og innovation kan øge eller mindske risikoen for omdirigering og misbrug
- (b) forstå de bestræbelser, som nedrustnings-, våbenkontrol- og ikkespredningsmiljøet allerede udfolder for at begrænse risikoen for omdirigering og misbrug af civil forskning og innovation, og
- (c) praktisere ansvarlig innovation på en måde, der begrænser risikoen for omdirigering og misbrug inden for udvikling og udbredelse af forskning og innovation.

3.1.1. Projektbeskrivelse

Dette projekt vil udarbejde tre særskilte sæt uddannelses- og kapacitetsopbygningsmaterialer.

- (a) Håndbog (1) – Håndbogen vil samle grundlæggende viden og midler, som AI-aktører kan anvende til i forsknings- og innovationsprocessen at vurdere og begrænse risikoen for uansvarlige aktørers omdirigering og misbrug af civil AI-teknologi. Den vil drøfte, hvorfor og hvordan beslutninger om udvikling og udbredelse af forskning og innovation kan indvirke på risikoen for omdirigering og misbrug og derved skabe trusler mod eller muligheder for fred og sikkerhed. Den vil også indføre relevante folkeretlige forpligtelser og eksportkontrolforpligtelser ud over de sikkerheds- og sikringshensyn, der drøftes i militær- samt nedrustnings-, våbenkontrol- og ikkespredningskredse, og give eksempler på processer og redskaber til at praktisere ansvarlig innovation, f.eks. metoder til vurdering af teknologisk indvirkning og modeller for risikovurdering.

- (b) *Podcastserie (~ 10)* – Disse podcasts vil fungere som et tilgængeligt og engagerende medium for AI-aktører, så de kan lære, hvorfor og hvordan ansvarlige AI-innovationsprocesser kan støtte international fred og sikkerhed gennem begrænsning af de risici, der er forbundet med uansvarlige aktørers omdirigering og misbrug. Serien vil gennemgå vigtige emner (f.eks. mønstret for omdirigering og misbrug af forskning og innovation med dobbelt/generel anvendelse, de humanitære, strategiske og politiske udfordringer i forbindelse med potentielt misbrug af civil AI-forskningsinnovation, de udfordringer, som nedrustnings-, våbenkontrol- og ikke-spredningskredse står over for i forbindelse med risikobegrænsningsindsatsen, hvordan ansvarlig innovation kan praktiseres gennem risikovurdering, overholdelse af eksportkontrol, indbygget risikobegrænsning, ansvarlig offentliggørelse, kendskab til kunder og erfaring med skrivebordsøvelser) og vil blive struktureret omkring interview, som projektgruppen vil gennemføre med repræsentanter fra relevante miljøer.
- (c) *Blogserie (9-10)* – Gruppen vil udarbejde en kurateret serie af blogindlæg, der har til formål at øge synligheden af de bestræbelser, der forsøger at overskride grænserne mellem de civilt fokuserede "ansvarlig AI"- og våbenkontrol- og ikke-spredningsmiljøer. Blogserien vil udgøre en platform for udbredelse af indsigter, idéer og løsninger vedrørende identifikation og håndtering af risici i forbindelse med omdirigering og misbrug af civil AI i forsknings- og innovationsprocessen. Bloggen vil søge at repræsentere mangfoldigheden af tanker og perspektiver i AI-sektoren.

Dette materiale vil blive udbredt offentligt via gennemførelsesaktørernes websteder og deres tilstedeværelse i de sociale medier og gennem direkte kommunikation med relevante akademiske enheder, civile faglige AI-sammenslutninger og andre passende grupper.

3.1.2. Forventede projektresultater

Dette projekt forventes at udarbejde et nyt sæt materialer, hvormed civile AI-aktører kan blive bevidste om, a) hvordan civil AI-forskning og -innovation kan omdirigeres og misbruges på måder, som kan udgøre trusler mod international fred og sikkerhed, b) hvordan sådanne trusler håndteres af nedrustnings-, våbenkontrol- og ikke-spredningsmiljøet, og c) hvordan AI-aktører kan bidrage yderligere til at begrænse sådanne trusler gennem ansvarlige innovationsprocesser.

Dette forventes at fremme den civile AI-sektors inddragelse i at begrænse den risici trussel, som omdirigering og misbrug af civil AI kan udgøre mod international fred og sikkerhed, forbedre tekniske aktørers kapacitet til at deltage i relevante processer i nedrustnings-, våbenkontrol- og ikke-spredningsmiljøet samt støtte inddragelse af nye målgrupper, der ikke traditionelt indgår i uddannelsesindsatsen inden for nedrustning og ikke-spredning.

Materialet forventes også at støtte gennemførelsen af de andre projekter og vil tjene som grundlag for uddannelses- og kapacitetsopbygningsaktiviteterne i projekt 2 samt dialog- og inddragelsesaktiviteterne i projekt 3. Disse aktiviteter forventes på deres side at indgå i udarbejdelsen og videreudviklingen af materialet. En sådan iterativ tilgang forventes at bidrage til at tackle potentielle hindringer for deres fremme, udbredelse og anvendelse i AI-miljøet, herunder spørgsmål vedrørende sprog, indhold, kontekst og tilgængelighed, hvilket kan udelukke deres indvirkning på verdensplan, navnlig i det globale Syd.

3.2. Projekt 2 – Uddannelses- og kapacitetsopbygningsaktiviteter for fremtidige AI-aktører

3.2.1. Projektformål

Formålet med projekt 2 er at støtte integration af problemet med uansvarlige aktørers omdirigering og misbrug af civil AI-forskning i uddannelsen af fremtidige generationer af AI-aktører. På lang sigt vil dette sikre, at studerende inden for naturvidenskab, teknologi, ingeniørvirksomhed og matematik (STEM), der præger fremtiden for AI, er opmærksomme på de negative indvirkninger, som uansvarlige aktørers omdirigering og misbrug af deres arbejde kan have på international fred og sikkerhed, og at de har de nødvendige grundlæggende redskaber til at identificere og begrænse en sådan risiko i forsknings- og innovationsprocessen.

Dette projekt vil gennemføre en række uddannelses- og kapacitetsopbygningsworkshopper med undervisere og studerende i samarbejde med udvalgte internationale universiteter og aktører fra erhvervslivet. Projektet søger derved at udvikle kapacitetsopbyggende aktiviteter, som undervisere og udarbejdere af akademiske læseplaner kan anvende til at medtage i uddannelsen af fremtidige AI-aktører (f.eks. kurser om AI-etik og ansvarlig AI-innovation), og overvejelser om risiciene for uansvarlige aktørers omdirigering og misbrug af civil AI-forskning og -innovation og sammenkæde disse med den bredere freds- og sikkerhedskontekst. Gennem disse workshopper vil projektet også søge at identificere et netværk af interesserede undervisere, udarbejdere af læseplaner og studerende, som vil støtte udbredelse og fremme af projektuddannelsesmateriale og kapacitetsopbygningsaktiviteter i AI-uddannelsessamfundet og AI-aktørsamfundet. Denne netværkskomponent har til formål at sikre projekternes bæredygtighed ud over deres umiddelbare varighed og at gøre det muligt at opbygge stærkere forbindelser til støtte for civil teknisk inddragelse i retning af større freds-, sikkerheds-, nedrustnings- og våbenkontrollsmål.

3.2.2. Projektbeskrivelse

Dette projekt vil gennemføre en række uddannelses- og kapacitetsopbygningsworkshopper med undervisere og studerende fra udvalgte universiteter i hele verden. De vil bestå af en blanding af forelæsninger og interaktive aktiviteter, som vil give undervisere og studerende muligheder for at reflektere over, hvordan civil AI-forskning og -innovation kan omdirigeres og misbruges på måder, der kan udgøre en trussel mod international fred og sikkerhed, og hvordan sådanne trusler kan identificeres, forebygges eller begrænses i forsknings- og innovationsprocessen eller gennem andre styringsprocesser. Disse aktiviteter vil bygge på tidligere mindre pilotarbejde udført af UNODA, som eksperimenterede med metoder til at inddrage og bevidstgøre STEM-studerende om betydningen af at overveje den bredere indvirkning af deres arbejde samt inddrage ekspertise uden for deres hovedområder. Konkret vil aktiviteterne bestå af

- (a) regionale kapacitetsopbyggende workshopper for undervisere og studerende (4) – De regionale workshopper vil gennemføre og fremme aktiviteter, som undervisere kan anvende til at opbygge STEM-studerendes kapacitet inden for ansvarlig AI-innovation

med særligt fokus på, hvordan risikoen for uansvarlige aktørers omdirigering og misbrug af civil AI-teknologi vurderes og begrænses. Hver workshop vil blive tilrettelagt med et EU-universitet og et højtprofileret universitet fra en anden region i verden, hvorved en blandet gruppe deltagere fra EU altid knyttes sammen med en blandet gruppe fra uden for EU. Workshopperne vil således omfatte Latinamerika og Caribien, Nordamerika, Afrika samt Asien og Stillehavsområdet. Dette vil gøre det muligt for studerende (på kandidat- og ph.d.-niveau) fra hele verden, herunder det globale Syd, at deltage. Workshoppen vil primært blive gennemført på engelsk, men hvor det er muligt, vil deltagerne få mulighed for at deltage i aktiviteter på grundlag af andre sproggrupper, og

- (b) international workshop om bæredygtig kapacitetsopbygning (1) – Workshoppen vil trække på erfaringerne fra de regionale workshopper og lette udveksling af oplysninger og erfaringer mellem undervisere og udvalgte studerende fra de universiteter, der deltager i projektet. Workshoppen vil drøfte, hvordan de aktiviteter og redskaber, der er udarbejdet i løbet af projektet, kan videreudvikles, og udbrede dem til andre end gruppen af deltagende universiteter. Den vil også drøfte, hvordan man kan støtte studerendes inddragelse i ansvarlig AI, der imødegår omdirigerings- og misbrugsrisici for den internationale fred og sikkerhed, når de er kommet ind i arbejdsstyrken.

SIPRI's og UNODA's netværk og tilstedeværelse i Afrika, Asien og Stillehavsområdet, Europa samt Nord- og Sydamerika vil blive anvendt til at lette og støtte aspekter af aktiviteterne, hvor det er relevant.

3.2.3. Forventede projektresultater

Projektet forventes at skabe modeller for kapacitetsopbygnings- og inddragelsesaktiviteter, som undervisere og udarbejdere af akademiske læseplaner kan kopiere for at bevidstgøre fremtidige AI-aktører om problemerne med uansvarlige aktørers omdirigering og misbrug af civil AI, og hvordan de kan bidrage til at begrænse disse problemer gennem ansvarlige innovationsprocesser. Efter afslutningen af projektaktiviteterne forventes deltagerne (undervisere, men også STEM-studerende) at kunne anvende og fremme ansvarlige innovationsredskaber, -metoder og -koncepter til at identificere og begrænse risikoen for omdirigering og misbrug inden for udvikling og udbredelse af civil AI-forskning og -innovation.

Projektaktiviteterne forventes desuden at skabe et netværk af undervisere, udarbejdere af læseplaner og studerende, som ikke blot vil fremme projektaktiviteterne inden for AI-uddannelsesmiljøer og -fagmiljøer (f.eks. på konferencer i IEEE Computational Intelligence Society), men også stå til rådighed for at bidrage med teknisk kapacitet til statsledede internationale styringsprocesser (f.eks. konventionen om visse konventionelle våbens proces for fremspirende teknologier på området dødbringende automatvåbensystemer).

Disse aktiviteters kort- og langsigtede værdi vil blive påvist gennem undersøgelser før og efter aktiviteterne.

3.3. Projekt 3 – Fremme af længeresigtet bæredygtig udvikling, udbredelse og indvirkning af ansvarlig innovation inden for AI med henblik på fred og sikkerhed

3.3.1. Projektformål

Formålet med projekt 3 er at lette længeresigtet bæredygtig udvikling, udbredelse og indvirkning af ansvarlig innovation inden for AI som et middel til at begrænse den trussel, som omdirigering og misbrug af civil AI-forskning og -innovation kan udgøre for fred og sikkerhed. Det søger at gøre dette gennem rundbordsdrøftelser med AI-branchen, multiinteressentdialoger, udarbejdelse af en offentlig rapport og målrettede udbredelsesaktiviteter. Projektet søger at sikre, at det arbejde, der udføres, navnlig uddannelses-, kapacitetsopbygnings- og inddragelsesaktiviteterne, når ud til og indvirker på AI-miljøet generelt på alle niveauer (fra studerende til ingeniører og andre AI-aktører) og på tværs af geografiske, sektorspecifikke og andre grænser. For at øge muligheden for at opnå en bred og dybtgående indvirkning er det væsentligt at samarbejde med faglige organisationer på dette område, såsom IEEE, og gennemføre flerdimensionel inddragelse på tværs af akademiske, industrielle og andre siloer. En sådan indsats vil give interesserede repræsentanter fra forskellige AI-miljøer mulighed for at tage ejerskab til problemet og udtrykke deres egne synspunkter om, hvordan risikobegrænsningsindsatsen kan gennemføres og fremmes bæredygtigt i og på tværs af det globale AI-miljø. Det er også vigtigt for projektets langsigtede værdi for stater, mellemstatslige organisationer og andre, at AI-aktører kan lære af og samarbejde med regeringseksperter, der beskæftiger sig med risikobegrænsning i forbindelse med nedrustning, våbenkontrol og ikkespredning. Det er endvidere afgørende for bæredygtigheden at sikre, at de indsigter, der skabes gennem inddragelsesaktiviteterne, analyseres, konsolideres og udbredes på passende vis.

3.3.2. Projektbeskrivelse

Dette projekt består af følgende hovedelementer:

- (a) multiinteressentdialoger om "ansvarlig AI-innovation med henblik på fred og sikkerhed" (op til 9) – Denne serie virtuelle dialogmøder vil samle eksperter fra den akademiske verden, forskning, den private sektor og traditionel våbenkontrol fra og uden for EU for at drøfte
- i. teknologiske tendenser, der kan skabe udbredelses-, omdirigerings- og misbrugsrisici med indvirkninger på international fred og sikkerhed
 - ii. hvordan man kan beskæftige sig med risikobegrænsning gennem ansvarlige innovationsprocesser, -metoder og -midler, samt muligheder og udfordringer for dialog og videndeling mellem interessentmiljøer, herunder dem, der arbejder i andre sektorer, såsom den biologiske og den kemiske branche, og
 - iii. den potentielle værdi af, formålet med og formatet af et selvbærende netværk af eksperter og dialogaktiviteter. Ekspertgruppen vil mødes flere gange om året og arbejde på at tilrettelægge to offentlige arrangementer for det bredere samfund.

Af de ni virtuelle møder er det hensigten, at to skal være åbne for offentligheden for at lette en bredere høring

- (b) rundbordsdrøftelser i den private sektor (op til 6) – Denne serie virtuelle rundbordsdrøftelser vil indlede en dialog med aktører, der arbejder med ansvarlige AI-innovationsprocesser i den private sektor (f.eks. partnerskabet om AI), om, hvordan de kan bidrage til at minimere risikoen for uansvarlige aktørers omdirigering og misbrug af civile AI-teknologier, samt undersøge mulige incitamenters indvirkning på udvikling i den private sektor til at gøre dette. Emnerne vil omfatte
- relevansen af den internationale sikkerheds- og nedrustningskontekst for den private sektor
 - de(t) retlige miljø(er), hvori udvikling, indsættelse og anvendelse af AI findes i verden
 - hvordan der kan bygges på risikovurderingsmekanismer og andre foranstaltninger, som er en del af eller kan integreres i ansvarlige innovationsprocesser og virksomheders overholdelsesprogrammer, og
 - erfaringer, der skal indhøstes fra andre brancher, processer og rammer vedrørende våbenkontrol (f.eks. den biologiske og den kemiske branche)
- (c) rapport om AI-miljøperspektiver vedrørende våbenkontrol og risikobegrænsning inden for AI, der er rettet mod AI- og våbenkontrollmiljøerne (1) – Udarbejdelsen af denne rapport vil opsamle og konsolidere resultaterne og anbefalingerne fra projektet i et enkelt referencedokument, som er rettet mod både det civilt ansvarlige AI-miljø og våbenkontrollmiljøerne. Rapporten vil drøfte, hvordan internationale trusler mod fred og sikkerhed forbundet med omdirigering og misbrug af civil AI-forskning og -innovation kan identificeres, vurderes og imødegås
- (d) udbredelsesarrangementer rettet mod høring af og samarbejde med AI- og våbenkontrollmiljøerne (skal fastlægges) – Gruppen vil søge muligheder for at formidle arbejdet og dets resultater og anbefalinger gennem hele projektet. Arrangementernes format og præsentationens indhold vil blive skræddersyet til målgruppens behov. De kan omfatte møder i CONOP, Den Europæiske Alliance vedrørende Kunstig Intelligens' forsamling, Gruppen af Regeringseksperters vedrørende Fremspirende Teknologier på Området Dødbringende Automatvåbensystemer, Den Tværinstitutionelle Arbejdsgruppe vedrørende Kunstig Intelligens (IAWG-AI), ITU's initiativ "AI for Good", UNIDIR's årlige innovationsdialog og Institute of Electrical and Electronics Engineers. Gruppen vil også søge at samarbejde bilateralt med relevante interessenter fra regeringen, den akademiske verden og den private sektor.

3.3.3. Forventede projektresultater

Dette projekt forventes at skabe grundlag for bæredygtig udvikling, udbredelse og indvirkning af ansvarlig innovation inden for AI-processer, der imødegår udbredelses-, omdirigerings- og misbrugsrisici og deres konsekvenser for freden og sikkerheden ud over den umiddelbare varighed af Rådets afgørelse.

Multiinteressentdialogen forventes at udgøre en model for informationsudveksling og samarbejde om risikobegrænsning, ikke kun inden for det globale AI-miljø, men også mellem

det civilt ansvarlige AI-miljø og nedrustnings-, våbenkontrol- og ikkespredningsmiljøerne. En sådan model kan anvendes til at gøre de politiske beslutningstagere fortrolige med vigtige teknologiske og videnskabelige fremskridt, der er relevante for ansvarlig innovation inden for AI, og desuden gøre de tekniske målgrupper fortrolige med det miljø, som de politiske beslutningstagere i øjeblikket arbejder i. Projektet forventes at lette bæredygtige forbindelser og bæredygtigt engagement mellem interesserede aktører i og på tværs af disse forskellige miljøer. Sådanne forskelligartede netværksvirkninger forventes at muliggøre en større udvikling og udstrakt fremme af ansvarlig innovation inden for AI med henblik på fred og sikkerhed ud over projektets tidsramme.

Dialogen i den private sektor forventes at muliggøre en større og dybere inddragelse af den private AI-sektor i identifikation, forebyggelse og begrænsning af de trusler mod fred og sikkerhed, der stammer fra omdirigering og misbrug af civil AI-forskning og -innovation. Projektet forventes at give nøgleaktører inden for processer i den private sektor større forståelse for og ejerskab til de problemer, som det søger at løse. Desuden har det til formål at lette en (bredere) antagelse og gennemførelse af ansvarlige innovationsprocesser, -metoder og -midler i eksisterende virksomhedsrisikostyringsmekanismer og -procedurer.

Multiinteressentdialogen og rundbordsdrøftelsen i den private sektor forventes endvidere at skabe indsigt i en række væsentlige spørgsmål, herunder a) hvordan ansvarlige innovationsmetoder og -midler kan videreudvikles og anvendes yderligere med henblik på at identificere, forebygge og begrænse de risici, der er forbundet med omdirigering og misbrug af civil AI-forskning og -innovation, b) hvordan AI-forskning og -innovation kan anvendes positivt til at støtte freds- og sikkerhedsmål (f.eks. applikationer til tidlig varsling af konflikter og humanitær bistand), og c) hvordan en større dialog og informationsudveksling kan lettes mellem risikobegrænsningsbestræbelserne i det civilt ansvarlige AI-miljø (dvs. forskellige IEEE-ledede initiativer) og de bestræbelser, der allerede udfoldes i nedrustnings-, våbenkontrol- og ikkespredningsmiljøet på mellemstatsligt plan.

Rapport- og udbredelsesaktiviteterne vil analysere, konsolidere og udbrede de indsigter, der er opnået gennem projekt 1, 2 og 3, og dermed støtte fremme af projektaktiviteternes resultater og anbefalinger i det globale AI-miljø og i det politiske samfund. De forventes også at bidrage til at sikre bæredygtigheden af indvirkningen ud over projekternes tidsramme.

4. VARIGHED

Den samlede anslåede varighed af projekternes gennemførelse er 36 måneder.