

Bruxelles, 26. listopada 2022.
(OR. en)

14095/22
ADD 1

LIMITE

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

PRIJEDLOG

Od:	Visoki predstavnik Unije za vanjske poslove i sigurnosnu politiku, potpisao glavni tajnik Stefano SANNINO
Datum primitka:	26. listopada 2022.
Za:	Thérèse BLANCHET, glavna tajnica Vijeća Europske unije
Predmet:	Prijedlog Visokog predstavnika Unije za vanjske poslove i sigurnosnu politiku Vijeću za odluku Vijeća o potpori provedbi projekta „Promicanje odgovornih inovacija u području umjetne inteligencije za mir i sigurnost”

Za delegacije se u prilogu nalazi dokument HR(2022) 238.

Priloženo: HR(2022) 238

EUROPSKA SLUŽBA ZA VANJSKO DJELOVANJE



PROJEKTNI DOKUMENT

Promicanje odgovornih inovacija u području umjetne inteligencije za mir i sigurnost

HR(2022)238

1. KONTEKST

Nedavni napredak u području umjetne inteligencije pruža nove mogućnosti za potporu i očuvanje mira i sigurnosti, primjerice tehnološkim poboljšanjima u područjima kao što su rano upozoravanje na sukobe i provjera kontrole oružja i izvoza. S druge strane, taj je napredak omogućio i nove načine za generiranje ili pogoršanje napetosti, sukoba i nesigurnosti među državama i unutar njih. Rizici koji proizlaze iz određenih primjena umjetne inteligencije, kao što su smrtonosni autonomni oružani sustavi, postali su novi veliki problem zajednice za kontrolu oružja. Faktor rizika koji zaslužuje veću pozornost i u odnosu na koji trenutačni diplomatski odgovori i odgovori u području kontrole oružja možda nisu dostatni jest preusmjeravanje i zlouporaba civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera, uključujući zlonamjerne nadržavne aktere, kao što je zlouporaba generativnih suparničkih mreža (engl. *Generative Adversarial Networks*, GAN) za izradu uvjerljivih krivotvorenih sadržaja (engl. *deepfakes*) za kampanje dezinformiranja.

Umjetna inteligencija razvojna je tehnologija s velikim potencijalom opće upotrebe. Određeni akteri mogli bi (relativno jednostavno) pristupiti istraživanjima i inovacijama u području umjetne inteligencije koji se razvijaju za civilnu primjenu i prenamijeniti ih za štetnu ili disruptivnu namjenu koja bi mogla utjecati na međunarodni mir i sigurnost. Preusmjeravanje i zlouporaba civilne tehnologije nisu nove pojave niti su svojstvene samo umjetnoj inteligenciji. To je nedavno zabilježeno i u povezanom području robotike, kada je Daiš/ISIS u Siriji pretvarao rekreativne bespilotne letjelice u oružje te ih upotrebljavao. Međutim, kod umjetne inteligencije problem je složeniji zbog više čimbenika: neopipljive i izrazito promjenjive prirode algoritama i podataka u području umjetne inteligencije, zbog čega je teško kontrolirati njihov prijenos/širenje; vodeće uloge privatnog sektora u ekosustavu istraživanja, razvoja i inovacija te s tim povezane potrebe da se algoritmi u njihovu vlasništvu zaštite; te globalne dostupnosti ljudskog stručnog znanja i materijalnih resursa kojima se tehnologije umjetne inteligencije mogu prenamijeniti. Istodobno, osobe koje rade u području umjetne inteligencije u civilnom sektoru prečesto nisu svjesne mogućih posljedica koje bi preusmjeravanje i zlouporaba njihova rada mogli imati za međunarodni mir i

sigurnost ili oklijevaju sudjelovati u aktualnim raspravama u okviru zajednice za kontrolu i neširenje oružja o rizicima povezanim s umjetnom inteligencijom.

Potrebno je poduprijeti veći angažman civilne zajednice u području umjetne inteligencije u razumijevanju i ublažavanju rizika za mir i sigurnost koji su povezani s preusmjeravanjem i zlouporabom civilne tehnologije umjetne inteligencije od strane neodgovornih aktera. Međunarodni institut za mirovna istraživanja u Stockholmu (SIPRI) i Ured Ujedinjenih naroda za poslove razoružanja (UNODA) smatraju da bi se to moglo postići većim promicanjem odgovornih inovacija u globalnoj civilnoj zajednici u području umjetne inteligencije. Dosadašnji rad SIPRI-ja i UNODA-e pokazao je da bi odgovorne inovacije, kao mehanizam samoupravljanja, mogle globalnoj civilnoj zajednici u području umjetne inteligencije pružiti praktične alate i metode za utvrđivanje, sprečavanje i ublažavanje rizika koje bi preusmjeravanje i zlouporaba civilnih istraživanja i inovacija u području umjetne inteligencije mogli predstavljati za mir i sigurnost. U njihovu radu utvrđene su i metodologije i nekoliko aktualnih civilno usmjerenih inicijativa za odgovornu umjetnu inteligenciju koje bi se mogle nadograditi kako bi se civilna zajednica u području umjetne inteligencije senzibilizirala za pitanja kontrole i neširenja oružja, stručne rasprave i stajališta država o odgovornom razvoju, širenju i upotrebi umjetne inteligencije, kao i za iskustva koja bi se mogla steći iz rada usmjerenog na odgovornost u obrambenom sektoru.¹ Ključno je da je u tom prethodnom radu jasno utvrđeno da je suradnja sa studentima iz područja znanosti, tehnologije, inženjerstva i matematike (STEM), koji još uvijek upotrebljavaju umjetnu inteligenciju u obrazovnom formatu, od središnje važnosti za sve učinkovite napore u području odgovornih inovacija.

2. CILJEVI

Cilj je tih projekata poduprijeti veći angažman civilne zajednice u području umjetne inteligencije u ublažavanju rizika koje preusmjeravanje i zlouporaba civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera mogu predstavljati za međunarodni mir i sigurnost. To se nastoji postići prije svega boljim razumijevanjem toga kako odluke u vezi s razvojem i širenjem istraživanja i inovacija u području umjetne inteligencije mogu utjecati na rizike od preusmjeravanja i zlouporabe te stvoriti rizik ili prilike za mir i sigurnost, a nakon toga promicanjem odgovornih inovacijskih procesa, metoda i alata koji mogu pomoći u osiguravanju miroljubive primjene civilnih inovacija i odgovornog širenja znanja o umjetnoj inteligenciji. U tu će se svrhu projektima podupirati aktivnosti izgradnje kapaciteta, istraživanja i angažmana kojima će se *i.* povećati kapacitet globalne civilne zajednice u području umjetne inteligencije kako bi se procesima odgovornih inovacija obuhvatili i prevladali rizici za mir i sigurnost koje predstavljaju preusmjeravanje i zlouporaba civilne umjetne inteligencije od strane neodgovornih aktera; *i.* poboljšati povezanost između napora za ublažavanje rizika u vezi s odgovornom umjetnom inteligencijom u civilnoj sferi i napora koji se već poduzimaju u zajednici za razoružanje te kontrolu i neširenje oružja na međuvladinoj razini. Ključno je da se njima ne namjeravaju uspostaviti novi standardi, načela ili propisi niti se na drugi način namjerava intervenirati u

¹ Metodologije, primjerice, uključuju preporučene prakse Instituta inženjera elektrotehnike i elektronike (IEEE) za procjenu učinka autonomnih i inteligentnih sustava na dobrobit ljudi (IEEE Std 7010 – 2020.) i Popis za procjenu pouzdane umjetne inteligencije (ALTAI) Stručne skupine na visokoj razini za umjetnu inteligenciju. Među inicijativama su: globalna inicijativa IEEE-a za etiku autonomnih i inteligentnih sustava; partnerstvo za umjetnu inteligenciju; globalno partnerstvo za umjetnu inteligenciju.

područja u nadležnosti država. Umjesto toga, cilj je projekata razviti civilne napore u području odgovornih inovacija kako bi se obuhvatili rizici za mir i sigurnost koje predstavljaju preusmjeravanje i zlouporaba civilne umjetne inteligencije od strane neodgovornih aktera te omogućiti obrazovanje o postojećim relevantnim međuvladinim naporima.

Kako bi se učinkovito doprlo do civilne zajednice u području umjetne inteligencije i utjecalo na nju, u okviru projekata primjenjuje se trodjelni pristup te se nastoji:

- (a) surađivati s profesorima – surađivati s odabranim profesorima i kreatorima akademskih kurikuluma na izradi i promicanju obrazovnih materijala koji se mogu upotrijebiti kako bi se razmatranja o rizicima za mir i sigurnost koji proizlaze iz preusmjeravanja i zlouporabe civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera uvrstila u osposobljavanje budućih stručnjaka u području umjetne inteligencije (npr. u predmete o etici i odgovornim inovacijama u području umjetne inteligencije);
- (b) surađivati sa studentima – pokazati odabranim studentima iz područja znanosti, tehnologije, inženjerstva i matematike (STEM) iz cijelog svijeta kako se rizici za mir i sigurnost koje predstavljaju preusmjeravanje i zlouporaba civilnog razvoja umjetne inteligencije od strane neodgovornih aktera mogu utvrditi, spriječiti ili ublažiti u okviru procesa istraživanja i inovacija ili putem drugih procesa upravljanja; i
- (c) surađivati s industrijom umjetne inteligencije – surađivati sa stručnim udrugama i tijelima za normizaciju kao što je Institut inženjera elektrotehnike i elektronike (IEEE) kako bi se i. stručnjacima iz tehničkog područja omogućili prilagođeni obrazovni materijali i aktivnosti sudjelovanja; ii. poduprla pozitivna upotreba umjetne inteligencije za mir i sigurnost; i iii. olakšali dijalog i razmjena informacija između stručnjaka iz akademske zajednice, privatnog sektora i vlade o načinima ublažavanja rizika od preusmjeravanja i zlouporabe civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera.

Takav pristup omogućuje da projekti dopru do svih članova zajednice u području umjetne inteligencije, tj. ne samo do postojećih stručnjaka, već i budućih generacija. Njime se omogućuje i angažman akademske zajednice, industrije i ostalih skupina iz tog područja te podupire održivost budućih napora uspostavom mreža kojima se premošćuju te granice.

Projektima se također nastoji iskoristiti pokretačka snaga i iskustvo SIPRI-ja i UNODA-e kako bi se utjecalo na globalnu zajednicu u području umjetne inteligencije, a ne samo na dionike iz EU-a. Jedinstveni položaj SIPRI-ja i UNODA-e omogućuje im da dopru do aktera u području umjetne inteligencije iz Afrike, azijsko-pacifičke regije, Europe te Sjeverne i Južne Amerike i potpomognu njihovu međusobnu suradnju. Oba subjekta također imaju iskustvo rada u drugim područjima znanosti i tehnologije koja se suočavaju sa sličnim izazovima dvojne namjene i širenja, uključujući biotehnologiju. Projektima se također nastoje iskoristiti uvjeti u Europskoj uniji, kao što su: a) postojanje naprednih višedioničkih procesa u području odgovorne umjetne inteligencije; b) visoka razina angažmana i stručnosti u pogledu pitanja razoružanja te kontrole i neširenja oružja u EU-u; c) raznovrsnost veza EU-ovih akademskih i istraživačkih organizacija i organizacija privatnog sektora s drugim regijama, posebno globalnim jugom, što će također biti jedan od glavnih ciljeva angažmana; i d) raznolikost nacionalnosti studenata, profesora i inženjera na sveučilištima, u istraživačkim ustanovama i privatnom sektoru.

Uključenost će biti jedna od ključnih sastavnica provedbe projektnih aktivnosti. Kako bi se učinkovito pružila potpora zajednici u području umjetne inteligencije, u projektima se prepoznaje da se ta zajednica sastoji od različitih aktera, a posebno i sljedeće:

- (a) rod je vrlo važan čimbenik. Stoga će pitanje roda biti integrirano u skladu sa strategijama UN-a o rodno osviještenoj politici i rodnom paritetu. Poticati će se i zahtijevati sudjelovanje žena u svim aktivnostima u okviru projekta; i
- (b) uključivanje osoba s invaliditetom i razumna prilagodba provodit će se tijekom cijelog trajanja projekta. To će podrazumijevati uklanjanje prepreka sudjelovanju osoba s invaliditetom te poduzimanje koraka za suradnju s osobama s invaliditetom i za osiguravanje toga da su njihovi stavovi i iskustva zastupljeni.

3. PROJEKTI

Tri projekta opisana u nastavku nadopunjavat će se i uzajamno podupirati, a elementi će se provoditi tijekom 36 mjeseci.

Projekt 1 – Izrada obrazovnih materijala i materijala za izgradnju kapaciteta civilne zajednice u području umjetne inteligencije

3.1.Svrha projekta

Projekt 1 usmjeren je na pružanje znanja i sredstava civilnim akterima u području umjetne inteligencije kako bi mogli procijeniti i ublažiti rizike koje preusmjeravanje i zlouporaba civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera mogu predstavljati za međunarodni mir i sigurnost. Cilj je izraditi obrazovne materijale i materijale za izgradnju kapaciteta koji će stručnjacima u području umjetne inteligencije iz svih regija, razina i sektora (uključujući profesore i kreatore kurikuluma iz područja umjetne inteligencije, studente STEM-a te inženjere i istraživače iz tog područja u akademskoj zajednici i privatnom sektoru) pružiti informacije i alate potrebne za

- (a) razumijevanje kako bi se civilna istraživanja i inovacije u području umjetne inteligencije mogli preusmjeriti i zloupotrijebiti na načine koji bi mogli predstavljati rizike za međunarodni mir i sigurnost te kako se odlukama u vezi s razvojem i širenjem istraživanja i inovacija ti rizici mogu povećati ili smanjiti;
- (b) razumijevanje napora koje je zajednica za razoružanje te kontrolu i neširenje oružja već uložila u ublažavanje rizika od preusmjeravanja i zlouporabe civilnih istraživanja i inovacija; i
- (c) provođenje odgovornih inovacija na način kojim se ublažava rizik od preusmjeravanja i zlouporabe u okviru razvoja i širenja istraživanja i inovacija.

3.1.1. Opis projekta

U okviru ovog projekta izradit će se tri odvojena skupa obrazovnih materijala i materijala za izgradnju kapaciteta.

- (a) Priručnik (1) – priručnik će sadržavati osnovno znanje i sredstva kako bi akteri u području umjetne inteligencije u procesu istraživanja i inovacija mogli procijeniti i ublažiti rizike od preusmjeravanja i zlouporabe civilne tehnologije umjetne inteligencije od strane neodgovornih aktera. U njemu će biti objašnjeno kako i zašto odluke u vezi s razvojem i širenjem istraživanja i inovacija mogu utjecati na rizike od preusmjeravanja i zlouporabe te dovesti do rizika ili prilika za mir i sigurnost. U priručniku će biti navedeni i relevantno međunarodno pravo i obveze u pogledu kontrole izvoza, kao i pitanja sigurnosti i zaštite o kojima se raspravlja u vojnim krugovima i zajednici za razoružanje te kontrolu i neširenje oružja te će biti predstavljeni primjeri procesa i alata za provedbu odgovornih inovacija, kao što su metodologije za procjenu tehnološkog učinka i predlošci za procjenu rizika.
- (b) Seriya podcasta (~10) – podcasti će biti pristupačan i zanimljiv medij putem kojeg će akteri u području umjetne inteligencije moći saznati kako i zašto se procesima odgovornih inovacija u području umjetne inteligencije mogu poduprijeti međunarodni mir i sigurnost ublažavanjem rizika koje predstavljaju preusmjeravanje i zlouporaba od strane neodgovornih aktera. U njima će se obrađivati važne teme (npr. obrazac preusmjeravanja i zlouporabe istraživanja i inovacija dvojne/opće namjene; humanitarni, strateški i politički izazovi povezani s potencijalnom zlouporabom civilnih istraživanja i inovacija u području umjetne inteligencije; izazovi s kojima se suočava zajednica za razoružanje te kontrolu i neširenje oružja u nastojanjima da se ublaže rizici; kako provoditi odgovorne inovacije na temelju procjene rizika; usklađenost kontrole izvoza; sustavno smanjenje rizika; odgovorno objavljivanje; poznavanje klijenata i iskustvo stožernih vježbi) te će biti strukturirani oko razgovora koje će projektni tim voditi s predstavnicima relevantnih zajednica.
- (c) Seriya objava na blogu (9 – 10) – tim će proizvesti seriju odabranih objava na blogu čiji je cilj povećati vidljivost napora kojima se pokušavaju premostiti granice između civilno usmjerene „odgovorne umjetne inteligencije” i zajednica za kontrolu i neširenje oružja. Blog će poslužiti kao platforma za širenje uvida, ideja i rješenja u pogledu utvrđivanja i prevladavanja rizika povezanih s preusmjeravanjem i zlouporabom civilne umjetne inteligencije u procesu istraživanja i inovacija. Nastojat će se prikazati različita mišljenja i perspektive iz sektora umjetne inteligencije.

Navedeni materijali bit će javno objavljeni na internetskim stranicama i društvenim mrežama aktera koji sudjeluju u provedbi te će se predstaviti u izravnoj komunikaciji s relevantnim akademskim subjektima, civilnim strukovnim udrugama u području umjetne inteligencije i drugim odgovarajućim skupinama.

3.1.2. Očekivani rezultati projekta

Očekuje se da će se ovim projektom izraditi novi skup materijala koji će civilne stručnjake u području umjetne inteligencije uputiti u to a) kako bi se civilna istraživanja i inovacije u području umjetne inteligencije mogli preusmjeriti i zloupotrijebiti na načine koji mogu predstavljati rizik za međunarodni mir i sigurnost, b) kako se ti rizici suzbijaju u okviru zajednice za razoružanje te kontrolu i neširenje oružja te c) kako bi stručnjaci u području umjetne inteligencije mogli dodatno doprinijeti ublažavanju takvih rizika procesima odgovornih inovacija.

Time bi se trebao povećati angažman civilnog sektora u području umjetne inteligencije u ublažavanju rizika koje preusmjeravanje i zlouporaba civilne umjetne inteligencije mogu

predstavljati za međunarodni mir i sigurnost; poboljšati kapacitet stručnjaka iz tehničkih područja za sudjelovanje u relevantnim procesima u okviru zajednice za razoružanje te kontrolu i neširenje oružja te podupirati angažman nove publike koja nije tradicionalno uključena u obrazovanje u području razoružanja i neširenja oružja.

Očekuje se da će se tim materijalima poduprijeti i provedba drugih projekata te da će oni poslužiti kao osnova za obrazovne aktivnosti i aktivnosti izgradnje kapaciteta u projektu 2 te za aktivnosti dijaloga i angažmana u projektu 3. Te bi aktivnosti, s druge strane, trebale doprinijeti izradi i poboljšanju materijala. Očekuje se da će takav iterativni pristup pomoći u otklanjanju mogućih prepreka promicanju, širenju i upotrebi materijala unutar zajednice za umjetnu inteligenciju, uključujući pitanja povezana s jezikom, sadržajem, kontekstom i dostupnošću, koje bi mogle onemogućiti njihov učinak na globalnoj razini, posebno na globalnom jugu.

3.2. Projekt 2 – Obrazovne aktivnosti i aktivnosti izgradnje kapaciteta za buduće stručnjake u području umjetne inteligencije

3.2.1. Svrha projekta

Svrha je projekta 2 poduprijeti uvrštavanje problema preusmjeravanja i zlouporabe civilnih istraživanja u području umjetne inteligencije od strane neodgovornih aktera u obrazovanje budućih generacija stručnjaka u području umjetne inteligencije. Time će se dugoročno osigurati da su studenti iz područja znanosti, tehnologije, inženjerstva i matematike (STEM), koji oblikuju budućnost umjetne inteligencije, svjesni negativnih učinaka koje bi preusmjeravanje i zlouporaba njihova rada od strane neodgovornih aktera mogli imati na međunarodni mir i sigurnost te da imaju osnovne alate potrebne za utvrđivanje i ublažavanje takvog rizika u procesu istraživanja i inovacija.

U okviru ovog projekta provest će se niz obrazovnih radionica i radionica za izgradnju kapaciteta s profesorima i studentima u suradnji s odabranim međunarodnim sveučilištima i akterima iz industrije. Projektom se stoga nastoje osmisliti aktivnosti izgradnje kapaciteta koje bi profesori i kreatori akademskih kurikuluma mogli iskoristiti i uključiti u osposobljavanje budućih stručnjaka u području umjetne inteligencije (npr. predmeti o etici i odgovornim inovacijama u području umjetne inteligencije) razmatranja o rizicima od preusmjeravanja i zlouporabe civilnih istraživanja i inovacija u području umjetne inteligencije od strane neodgovornih aktera te povezati te rizike sa širim kontekstom mira i sigurnosti. Tim će se radionicama u okviru projekta nastojati utvrditi i mreža zainteresiranih profesora, kreatora kurikuluma i studenata koji bi poduprli širenje i promicanje obrazovnih materijala i aktivnosti izgradnje kapaciteta iz projekta u obrazovnoj zajednici u području umjetne inteligencije i zajednici stručnjaka u tom području. Cilj je tog umrežavanja zajamčiti održivost projekata i nakon njihova neposrednog trajanja te omogućiti izgradnju snažnijih veza kojima će se poduprijeti angažman civilnih stručnjaka iz tehničkih područja kako bi se postigli veći ciljevi u pogledu mira, sigurnosti, razoružanja i kontrole oružja.

3.2.2. Opis projekta

U okviru ovog projekta provest će se niz obrazovnih radionica i radionica za izgradnju kapaciteta s profesorima i studentima s odabranih sveučilišta iz cijelog svijeta. Radionice će se sastojati od

predavanja i interaktivnih aktivnosti koje će profesorima i studentima pružiti priliku da razmisle o tome kako bi se civilna istraživanja i inovacije u području umjetne inteligencije mogli preusmjeriti i zloupotrijebiti na načine koji mogu predstavljati rizike za međunarodni mir i sigurnost te kako se takvi rizici mogu utvrditi, spriječiti ili ublažiti u okviru procesa istraživanja i inovacija ili drugih procesa upravljanja. Aktivnosti će se temeljiti na prethodnom manjem pilot-projektu koji je provela UNODA, a u kojem se eksperimentiralo s metodama za suradnju sa studentima iz područja STEM-a i njihovo senzibiliziranje o tome da je važno razmotriti širi učinak njihova rada te surađivati sa stručnjacima izvan njihovih matičnih područja. Konkretno, one će obuhvaćati:

- (a) regionalne radionice za izgradnju kapaciteta za profesore i studente (4) – u okviru regionalnih radionica provodit će se i promicati aktivnosti koje profesori mogu iskoristiti za izgradnju kapaciteta studenata iz područja STEM-a za odgovorne inovacije u području umjetne inteligencije, s posebnim naglaskom na tome kako procijeniti i ublažiti rizike od preusmjeravanja i zlouporabe civilne tehnologije umjetne inteligencije od strane neodgovornih aktera. Svaka radionica organizirat će se s jednim sveučilištem sa sjedištem u EU-u i jednim istaknutim sveučilištem iz neke druge svjetske regije, čime će se uvijek povezati raznolik skup sudionika iz EU-u s raznolikim skupom sudionika izvan EU-a. Radionice će obuhvatiti Latinsku Ameriku i Karibe, Sjevernu Ameriku, Afriku te Aziju i Pacifik. Time će se omogućiti sudjelovanje studenata (na diplomskoj i doktorskoj razini) iz cijelog svijeta, uključujući globalni jug. Radionice će se održavati prvenstveno na engleskom jeziku, ali ako to bude izvedivo, sudionici će imati priliku sudjelovati u aktivnostima koje se temelje na drugim jezičnim skupinama; i
- (b) međunarodnu radionicu o održivoj izgradnji kapaciteta (1) – radionica će se temeljiti na iskustvima stečenima na regionalnim radionicama i olakšati razmjenu informacija i iskustava među profesorima i odabranim studentima sa sveučilišta koja sudjeluju u projektu. Na radionici će se raspravljati o tome kako poboljšati aktivnosti i alate razrađene tijekom projekta i širiti ih izvan skupine sveučilišta koja sudjeluju u projektu. Raspravljat će se i o tome kako, nakon što studenti uđu na tržište rada, podupirati njihovo uključivanje u rad na odgovornoj umjetnoj inteligenciji kojom se prevladavaju rizici za međunarodni mir i sigurnost povezani s preusmjeravanjem i zlouporabom.

Mreže SIPRI-ja i UNODA-e i njihova prisutnost u Africi, azijsko-pacifičkoj regiji, Europi te Sjevernoj i Južnoj Americi iskoristit će se prema potrebi kako bi se olakšali i poduprli određeni aspekti aktivnosti.

3.2.3. Očekivani rezultati projekta

Očekuje se da će se projektom osmisliti modeli aktivnosti izgradnje kapaciteta i angažmana koje bi profesori i kreatori akademskih kurikuluma mogli replicirati kako bi se buduće stručnjake u području umjetne inteligencije senzibiliziralo o problemima preusmjeravanja i zlouporabe civilne umjetne inteligencije od strane neodgovornih aktera te o tome kako oni mogu pomoći u ublažavanju tih problema s pomoću procesa odgovornih inovacija. Očekuje se da će nakon završetka projektnih aktivnosti sudionici (profesori, ali i studenti iz područja STEM-a) moći upotrebljavati i promicati odgovorne inovacijske alate, metode i koncepte kako bi utvrdili i ublažili rizike od preusmjeravanja i zlouporabe u okviru razvoja i širenja civilnih istraživanja i inovacija u području umjetne inteligencije.

Očekuje se i da će se projektnim aktivnostima stvoriti mreža profesora, kreatora kurikuluma i studenata koji će promicati projektne aktivnosti u obrazovnoj i stručnoj zajednici u području umjetne inteligencije (npr. na konferencijama IEEE-ova Društva za računalnu inteligenciju), ali i biti u mogućnosti svojim tehničkim kapacitetima doprinijeti procesima međunarodnog upravljanja pod vodstvom države (npr. procesu Konvencije o određenom konvencionalnom oružju u vezi s tehnologijama u nastajanju u području smrtonosnih autonomnih oružanih sustava).

Kratkoročna i dugoročna vrijednost tih aktivnosti pokazat će se u anketama provedenima prije i poslije aktivnosti.

3.3. Projekt 3 – Olakšavanje dugoročnijeg održivog razvoja, širenja i učinka odgovornih inovacija u području umjetne inteligencije za mir i sigurnost

3.3.1. Svrha projekta

Svrha je projekta 3 olakšati dugoročniji održivi razvoj, širenje i učinak odgovornih inovacija u području umjetne inteligencije kako bi se ublažili rizici koje preusmjeravanje i zlorporaba civilnih istraživanja i inovacija u području umjetne inteligencije mogu predstavljati za mir i sigurnost. To se nastoji postići okruglim stolovima s predstavnicima industrije umjetne inteligencije, višedioničkim dijalozima, izradom javnog izvješća i ciljanim aktivnostima širenja informacija. Cilj je projekta osigurati da ostvareni rad, a posebno aktivnosti obrazovanja, izgradnje kapaciteta i angažmana, dopire do zajednice u području umjetne inteligencije u cjelini i utječe na nju, i to na svim razinama (od studenata do inženjera i drugih stručnjaka za umjetnu inteligenciju) i preko svih geografskih, sektorskih i drugih granica. Kako bi se povećala mogućnost ostvarivanja velikog i snažnog učinka, ključno je surađivati sa stručnim organizacijama iz tog područja, kao što je IEEE, i provoditi višedimenzionalne angažmane u akademskoj zajednici, industriji i ostalim skupinama iz tog područja. Takvim će se naporima zainteresiranim predstavnicima iz različitih zajednica u području umjetne inteligencije pružiti prilika da preuzmu odgovornost za taj problem i iznesu vlastita stajališta o tome kako se naponi za ublažavanje rizika mogu provoditi i promicati na održiv način unutar globalne zajednice u području umjetne inteligencije. Za osiguravanje dugoročne vrijednosti projekta državama, međuvladinim organizacijama i drugima važno je i da stručnjaci u području umjetne inteligencije mogu učiti od vladinih stručnjaka koji se bave ublažavanjem rizika u kontekstu razoružanja te kontrole i neširenja oružja te surađivati s njima. Kako bi se ostvarila održivost, ključno je osigurati da se uvidi dobiveni u okviru aktivnosti angažmana na odgovarajući način analiziraju, konsolidiraju i šire.

3.3.2. Opis projekta

Ovaj se projekt sastoji od sljedećih ključnih komponenti:

- (a) višedioničkih dijaloga o „odgovornim inovacijama u području umjetne inteligencije za mir i sigurnost” (do 9) – ova serija virtualnih sastanaka u obliku dijaloga okupit će stručnjake iz akademske zajednice, istraživanja, privatnog sektora i sektora tradicionalne kontrole oružja iz EU-a i šire kako bi raspravljali o:
 - i. tehnološkim trendovima koji mogu dovesti do rizika od širenja, preusmjeravanja i zlorporabe s posljedicama za međunarodni mir i sigurnost;

- ii. tome kako ublažavati rizike putem procesa, metoda i sredstava za odgovorne inovacije te prilika i izazova za dijalog i razmjenu znanja među zajednicama dionika, uključujući one koje djeluju u drugim sektorima, kao što su biološka i kemijska industrija; i
- iii. potencijalnoj vrijednosti, svrsi i formatu samoodržive mreže stručnjaka i aktivnosti dijaloga. Skupina stručnjaka sastajat će se nekoliko puta godišnje i raditi na organizaciji dvaju javnih događanja za širu zajednicu.

Od devet virtualnih sastanaka dva će biti otvorena za javnost kako bi se olakšalo šire savjetovanje;

(b) okruglih stolova s privatnim sektorom (do 6) – ovom će se serijom virtualnih okruglih stolova započeti dijalog s akterima koji rade s procesima odgovornih inovacija u području umjetne inteligencije u privatnom sektoru (npr. partnerstvo za umjetnu inteligenciju) o tome kako mogu doprinijeti smanjenju rizika od preusmjeravanja i zlouporabe civilnih tehnologija umjetne inteligencije od strane neodgovornih aktera, kao i istražiti mogući poticaji za to u okviru razvoja privatnog sektora. Teme će uključivati:

- i. važnost međunarodnog konteksta u području sigurnosti i razoružanja za privatni sektor;
- ii. pravno okružje za razvoj, uvođenje i upotrebu umjetne inteligencije u svijetu;
- iii. kako se osloniti na mehanizme za procjenu rizika i druge mjere koje su dio procesa odgovornih inovacija i programâ korporativne usklađenosti ili bi se mogli integrirati u njih; i
- iv. iskustva koja se mogu steći iz drugih industrija, procesa i okvira povezanih s kontrolom oružja (npr. biološke i kemijske industrije);

(c) izvješća o stajalištima zajednice u području umjetne inteligencije o kontroli oružja i ublažavanju rizika u području umjetne inteligencije, namijenjenog toj zajednici i zajednici za kontrolu oružja (1) – ovo će izvješće obuhvatiti nalaze i preporuke iz projekta, konsolidirane u jedinstveni referentni dokument namijenjen civilnoj zajednici za odgovornu umjetnu inteligenciju i zajednici za kontrolu oružja. U izvješću će se razmotriti kako utvrditi, procijeniti i prevladati rizike za međunarodni mir i sigurnost povezane s preusmjeravanjem i zlouporabom civilnih istraživanja i inovacija u području umjetne inteligencije;

(d) događanja u svrhu širenja informacija radi savjetovanja i suradnje sa zajednicom u području umjetne inteligencije i zajednicom za kontrolu oružja (još nije odlučeno) – tim će nastojati informirati o svojem radu, nalazima i preporukama tijekom cijelog projekta. Format događanja i sadržaj prezentacija bit će prilagođeni potrebama ciljnih skupina. Može se raditi o sastancima CONOP-a, skupštine Europskog saveza za umjetnu inteligenciju, skupine vladinih stručnjaka za nove tehnologije u području smrtonosnih autonomnih oružanih sustava, međuagencijske radne skupine za umjetnu inteligenciju (IAWG-AI), inicijative ITU-a „Umjetna inteligencija za opće dobro”, godišnjeg dijaloga UNIDIR-a o inovacijama i Instituta inženjera elektrotehnike i elektronike. Tim će također nastojati bilateralno surađivati s relevantnim dionicima iz vlade, akademske zajednice i privatnog sektora.

3.3.3. Očekivani rezultati projekta

Očekuje se da će se ovim projektom postaviti temelji za održivi razvoj, širenje i učinak procesa odgovornih inovacija u području umjetne inteligencije kojima se prevladavaju rizici od širenja, preusmjeravanja i zloporabe te njihove posljedice za mir i sigurnost nakon neposrednog trajanja odluke Vijeća.

Višedionički dijalog trebao bi poslužiti kao model za razmjenu informacija i suradnju u pogledu ublažavanja rizika ne samo unutar globalne zajednice u području umjetne inteligencije, već i između civilne zajednice za odgovornu umjetnu inteligenciju i zajednica za razoružanje te kontrolu i neširenje oružja. Takav bi se model mogao upotrijebiti za upoznavanje oblikovatelja politika s ključnim tehnološkim i znanstvenim napretkom važnim za odgovorne inovacije u području umjetne inteligencije, kao i za upoznavanje tehničkih stručnjaka s okruženjem u kojem oblikovatelji politika trenutačno djeluju. Očekuje se da će se projektom olakšati održivi odnosi i suradnja zainteresiranih aktera unutar tih različitih zajednica i među njima. Učinci tog heterogenog umrežavanja trebali bi omogućiti veći razvoj i široko promicanje odgovornih inovacija u području umjetne inteligencije za mir i sigurnost nakon završetka projekta.

Očekuje se da će se dijalogom s privatnim sektorom omogućiti veći i konkretniji angažman privatnog sektora u području umjetne inteligencije u utvrđivanju, sprečavanju i ublažavanju rizika za mir i sigurnost koji proizlaze iz preusmjeravanja i zloporabe civilnih istraživanja i inovacija u području umjetne inteligencije. Također se očekuje i da će projekt ključnim akterima u procesima privatnog sektora omogućiti bolje razumijevanje i preuzimanje odgovornosti za probleme koji se nastoje riješiti. Osim toga, cilj je projekta olakšati (široko) donošenje i provedbu procesa, metoda i sredstava za odgovorne inovacije u postojećim mehanizmima i postupcima za upravljanje korporativnim rizikom.

Očekuje se i da će se višedioničkim dijalogom i okruglim stolovima s privatnim sektorom dobiti uvid u niz bitnih pitanja, među kojima su i pitanja a) kako se metode i sredstva za odgovorne inovacije mogu dodatno poboljšati i primijeniti kako bi se utvrdili, spriječili i ublažili rizici povezani s preusmjeravanjem i zloporabom civilnih istraživanja i inovacija u području umjetne inteligencije; b) kako se istraživanja i inovacije u području umjetne inteligencije mogu pozitivno upotrijebiti za potporu ciljevima mira i sigurnosti (npr. primjene za rano upozoravanje na sukobe i humanitarnu pomoć); i c) kako poboljšati dijalog i razmjenu informacija između napora za ublažavanje rizika koje poduzima civilna zajednica za odgovornu umjetnu inteligenciju (tj. razne inicijative pod vodstvom IEEE-a) i napora koji su već u tijeku u zajednici za razoružanje te kontrolu i neširenje oružja na međuvladinoj razini.

U izvješću i aktivnostima širenja analizirat će se, konsolidirati i širiti saznanja stečena u okviru projekata 1, 2 i 3, čime će se pomoći u promicanju nalaza i preporuka iz projektnih aktivnosti unutar globalne zajednice u području umjetne inteligencije, kao i unutar zajednice za politike. Očekuje se i da će oni doprinijeti osiguravanju održivosti učinka i nakon završetka projekata.

4. TRAJANJE

Sveukupno procijenjeno trajanje provedbe projekata iznosi 36 mjeseci.

