Svet
Evropske unije

Bruselj, 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

PUBLIC

PREDLOG

Pošiljatelj:	za visokega predstavnika Unije za zunanje zadeve in varnostno politiko: generalni sekretar Stefano SANNINO
Datum prejema:	26. oktober 2022
Prejemnik:	Thérèse BLANCHET, generalna sekretarka Sveta Evropske unije
Zadeva:	Predlog visokega predstavnika Unije za zunanje zadeve in varnostno politiko Svetu za sklep Sveta v podporo izvajanju projekta „Spodbujanje odgovornih inovacij na področju umetne inteligence za mir in varnost“

Delegacije prejmejo priloženi dokument HR(2022) 238.

Priloga: HR(2022) 238

EVROPSKA SLUŽBA ZA ZUNANJE DELOVANJE

PUBLIC

PROJEKTNI DOKUMENT
Spodbujanje odgovornih inovacij na področju umetne inteligence za mir in varnost
HR(2022)238

1. OZADJE

Nedavni napredek na področju umetne inteligence je sprostil nove možnosti za podporo in ohranjanje miru in varnosti, na primer s tehnološkimi izboljšavami na področjih, kot so zgodnje opozarjanje na konflikte ter preverjanje orožja in nadzora izvoza. Po drugi strani je ta napredek omogočil nova sredstva za ustvarjanje ali poslabšanje napetosti, konfliktov in negotovosti med državami in znotraj njih. Tveganja, ki jih predstavljajo nekatere uporabe umetne inteligence, kot so smrtonosni avtonomni orožni sistem, so postala ena najbolj problematičnih vprašanj za skupnost za nadzor nad oborožitvijo. Eden od načinov tveganja, ki si zasluži več pozornosti in za katerega sedanji nadzor nad oborožitvijo in diplomatska prizadevanja morda ne zadoščajo za odzivanje nanj, je preusmerjanje in zloraba civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev, vključno z zlonamernimi nedržavnimi akterji, kot je zloraba generativnega kontradiktornega omrežja za izdelavo globokih ponaredkov za dezinformacijske kampanje.

Umetna inteligenca je omogočitvena tehnologija z velikim potencialom splošne uporabe. Nekateri akterji bi lahko do raziskav in inovacij na področju umetne inteligence, ki se razvijajo za civilne aplikacije, (razmeroma zlahka) dostopali in jim namenili nove namene za škodljive ali moteče uporabe, ki bi lahko vplivale na mednarodni mir in varnost. Preusmerjanje in zloraba civilne tehnologije nista nova pojava in nista edinstvena za umetno inteligenco. Na sorodnem področju robotike je bilo to nedavno ponazorjeno z oboroževanjem in uporabo rekreacijskih brezpilotnih zrakoplovov s strani Daiša/ISIS v Siriji. Vendar je v primeru umetne inteligence težava zapletena zaradi več dejavnikov: nematerialne in hitro spreminjajoče se narave algoritmov in podatkov umetne inteligence, zaradi česar je prenos/širjenje teh algoritmov in podatkov težko nadzorovati; vodilne vloge zasebnega sektorja v raziskovalnem, razvojnem in inovacijskem ekosistemu ter posledične potrebe po zaščiti lastniških algoritmov; ter globalne razpoložljivosti človeškega strokovnega znanja in materialnih virov, s katerimi je mogoče spremeniti namen tehnologij umetne inteligence. Medtem se tisti, ki delajo na področju umetne inteligence v civilnem sektorju, prepogosto ne zavedajo morebitnih posledic, ki bi jih lahko imela

preusmeritev in zloraba njihovega dela za mednarodni mir in varnost, ali pa se neradi udeležujejo obstoječih razprav o tveganjih, povezanih z umetno inteligenco v krogih nadzora nad oboroževanjem in neširjenja orožja.

Podpreti je treba večje sodelovanje civilne skupnosti na področju umetne inteligence pri razumevanju in zmanjševanju tveganj za mir in varnost, povezanih s preusmerjanjem in zlorabo civilne tehnologije umetne inteligence s strani neodgovornih akterjev. Za Mednarodni mirovni raziskovalni inštitut v Stockholmu (SIPRI) in Urad Združenih narodov za razorožitev (UN-ODA) bi bilo to mogoče doseči z večjim spodbujanjem odgovornih inovacij v svetovni civilni skupnosti na področju umetne inteligence. Preteklo delo SIPRI in UN-ODA je pokazalo, da bi lahko odgovorne inovacije kot mehanizem samouprave svetovni civilni skupnosti na področju umetne inteligence zagotovile praktična orodja in metode za opredelitev ter pomoč pri preprečevanju in zmanjševanju tveganj, ki bi jih za mir in varnost lahko pomenila preusmerjanje in zloraba civilnih raziskav in inovacij na področju umetne inteligence. V delu SIPRI in UN-ODA so bile opredeljene tudi metodologije in več tekočih civilno usmerjenih pobud o odgovorni umetni inteligenci, ki bi lahko temeljile na ozaveščanju civilne skupnosti na področju umetne inteligence o vprašanjih nadzora nad oboroževanjem in neširjenja orožja, strokovnih razpravah in državnih stališčih o odgovornem razvoju, razširjanju in uporabi umetne inteligence ter izkušnjah, pridobljenih pri delu na področju odgovornosti obrambnega sektorja.[1] Ključno je, da je bilo v tem predhodnem delu jasno opredeljeno, da je sodelovanje s študenti naravoslovja, tehnologije, inženirstva in matematike, ki se še vedno ukvarjajo z umetno inteligenco v izobraževalni obliki, osrednjega pomena za vsa učinkovita odgovorna inovacijska prizadevanja.

2. CILJI

Cilj teh projektov je podpreti večje sodelovanje civilne skupnosti na področju umetne inteligence pri zmanjševanju tveganj, ki jih lahko za mednarodni mir in varnost pomenita preusmerjanje in zloraba civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev. To naj bi dosegli, prvič, z boljšim razumevanjem, kako lahko odločitve pri razvoju in razširjanju raziskav in inovacij na področju umetne inteligence vplivajo na tveganja preusmerjanja in zlorabe ter posledično ustvarjajo tveganja ali priložnosti za mir in varnost, ter, drugič, s spodbujanjem odgovornih inovacijskih procesov, metod in orodij, ki lahko pomagajo zagotoviti mirno uporabo civilnih inovacij in odgovorno razširjanje znanja o umetni inteligenci. V ta namen podpirajo dejavnosti krepitve zmogljivosti, raziskav in sodelovanja, ki bodo *i)* prek odgovornih inovacijskih procesov okrepile zmogljivosti svetovne civilne skupnosti na področju umetne inteligence za vključevanje in obravnavanje tveganj za mir in varnost, ki jih predstavljata preusmerjanje in zloraba civilne umetne inteligence s strani neodgovornih akterjev; in *ii)* okrepile povezavo med prizadevanji za zmanjšanje tveganja na področju odgovorne umetne inteligence na civilnem področju in prizadevanji, ki že potekajo v skupnosti za razoroževanje,

[1] Metodologije vključujejo na primer priporočene prakse Inštituta inženirjev elektrotehnike in elektronike za ocenjevanje učinka avtonomnih in inteligentnih sistemov na dobro počutje ljudi (IEEE Std 7010–2020) in ocenjevalni seznam strokovne skupine na visoki ravni za umetno inteligenco za zaupanja vredno umetno inteligenco. Pobude vključujejo: globalno pobudo Inštituta inženirjev elektrotehnike in elektronike o etiki za avtonomne in inteligentne sisteme; partnerstvo za umetno inteligenco; svetovno partnerstvo za umetno inteligenco.

nadzor nad oboroževanjem in neširjenje orožja na medvladni ravni. Bistveno je, da ne nameravajo vzpostaviti novih standardov, načel ali predpisov ali kako drugače posegati na področja v pristojnosti držav. Namesto tega nameravajo razviti civilno odgovorna inovacijska prizadevanja za vključitev tveganj za mir in varnost, ki jih predstavljata preusmerjanje in zloraba civilne umetne inteligence s strani neodgovornih akterjev, ter zagotoviti izobraževanje o obstoječih ustreznih medvladnih prizadevanjih.

Da bi projekti učinkovito dosegli civilno skupnost na področju umetne inteligence in nanjo vplivali, uporabljajo tridelni pristop, katerega namen je:

(a) *sodelovanje z izobraževalci* – sodelovanje z izbranimi izobraževalci in razvijalci akademskih učnih načrtov pri razvoju in promociji učnega gradiva, ki se lahko uporabi za vključevanje upoštevanja mirovnih in varnostnih tveganj, ki izhajajo iz preusmerjanja in zlorabe civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev pri usposabljanju prihodnjih strokovnjakov na področju umetne inteligence (npr. v tečajih o etiki umetne inteligence in odgovornih inovacijah);

(b) *sodelovanje s študenti* – predstavitev izbranim študentom naravoslovja, tehnologije, inženirstva in matematike z vsega sveta, kako bi bilo mogoče prepoznati, preprečiti ali ublažiti tveganja za mir in varnost, ki jih predstavlja preusmerjanje in zloraba civilnega razvoja umetne inteligence s strani neodgovornih akterjev, v procesu raziskav in inovacij ali z drugimi procesi upravljanja; in

(c) *sodelovanje z industrijo umetne inteligence* – sodelovanje s strokovnimi združenji in organi za standardizacijo, kot je Inštitut inženirjev elektronike in elektronike, da bi i) tehničnim strokovnjakom razširjali prilagojena učna gradiva in dejavnosti vključevanja; ii) podpirali pozitivno uporabo umetne inteligence za mir in varnost; ter iii) olajšali dialog in izmenjavo informacij med strokovnjaki iz akademskih krogov, zasebnega sektorja in vlade o tem, kako bi lahko zmanjšali tveganje preusmerjanja in zlorabe civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev.

Tak pristop omogoča, da projekti dosežejo skupnost umetne inteligence na vseh ravneh, vključno s sedanjimi strokovnjaki, pa tudi prihodnjimi generacijami. Omogoča tudi sodelovanje med akademskimi krogi, industrijo in drugimi področji delovanja ter podpira trajnost prihodnjih prizadevanj z vzpostavitvijo mrež, ki presegajo te meje.

Namen projektov je tudi izkoristiti povezovalno moč in izkušnje SIPRI in UN-ODA, da bi vplivali na skupnost umetne inteligence na svetovni ravni, ne le na deležnike EU. SIPRI in UN-ODA sta v edinstvenem položaju, da dosežeta in olajšata sodelovanje med akterji umetne inteligence iz Afrike, azijsko-pacifiške regije, Evrope ter Severne in Južne Amerike. Oba subjekta imata tudi izkušnje z delom na drugih področjih znanosti in tehnologije, ki se soočajo s podobnimi izzivi dvojne rabe in širjenja, vključno z biotehnologijo. Namen projektov je tudi izkoristiti razmere v Evropski uniji, kot so a) napredni postopki z več deležniki na področju odgovorne umetne inteligence; b) visoka stopnja sodelovanja in strokovnega znanja na področju razoroževanja, nadzora nad oboroževanjem in neširjenja orožja v EU; c) raznolikost povezav, ki jih imajo akademske in raziskovalne organizacije ter organizacije zasebnega sektorja v EU z drugimi regijami, zlasti na svetovnem jugu, kar bo tudi glavni cilj sodelovanja; ter d) raznolikost

narodnosti študentov, izobraževalcev in inženirjev na univerzah, raziskovalnih ustanovah in v zasebnem sektorju.

Vključevanje bo osrednje vodilo pri izvajanju dejavnosti projektov. Za učinkovito podporo skupnosti umetne inteligence projekti priznavajo, da je skupnost umetne inteligence sestavljena iz različnih akterjev, zlasti da

(a) je spol zelo pomemben dejavnik. Zato bo vidik spola vključen v skladu s strategijami vključevanja vidika spola in enakosti spolov v celotnem sistemu ZN. Spodbujalo in zahtevalo se bo sodelovanje žensk v vseh dejavnostih v okviru projekta; in

(b) povsod se bosta izvajala vključevanje invalidov in razumna prilagoditev potrebam. To bo vključevalo odpravo ovir za udeležbo invalidov ter zagotavljanje, da se sprejmejo ukrepi za sodelovanje z invalidi in olajšanje zastopanja njihovih vsebinskih stališč in izkušenj.

3. PROJEKTI

Trije projekti, opisani v nadaljevanju, naj bi se dopolnjevali in vzajemno podpirali, elementi pa naj bi trajali 36 mesecev.

Projekt 1 – Priprava učnega gradiva in gradiva za krepitev zmogljivosti za civilno skupnost na področju umetne inteligence

3.1.Namen projekta

Projekt 1 se osredotoča na zagotavljanje znanja in sredstev za civilne akterje na področju umetne inteligence pri ocenjevanju in zmanjševanju tveganj, ki jih lahko za mednarodni mir in varnost pomenita preusmerjanje in zloraba civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev. Njegov cilj je pripraviti učno gradivo in gradivo za krepitev zmogljivosti, ki bo strokovnjakom na področju umetne inteligence iz vseh regij, ravni in sektorjev (vključno z izobraževalci, ki se osredotočajo na umetno inteligenco, razvijalci učnih načrtov, študenti na področju naravoslovja, tehnologije, inženirstva in matematike ter inženirji in raziskovalci na področju umetne inteligence v akademskem in zasebnem sektorju) zagotovilo informacije in orodja, potrebna za:

(a) razumevanje, kako bi bilo mogoče civilne raziskave in inovacije na področju umetne inteligence preusmeriti in zlorabiti na način, ki bi lahko pomenil tveganje za mednarodni mir in varnost, ter kako lahko odločitve pri razvoju in razširjanju raziskav in inovacij povečajo ali zmanjšajo tveganje preusmerjanja in zlorabe;

(b) razumevanje prizadevanj, ki jih že izvaja skupnost za razoroževanje, nadzor nad oborožitvijo in neširjenje orožja za zmanjšanje tveganja preusmerjanja in zlorabe civilnih raziskav in inovacij; in

(c) izvajanje odgovornih inovacij na način, ki zmanjšuje tveganje preusmerjanja in zlorabe pri razvoju in razširjanju raziskav in inovacij.

3.1.1. Opis projekta

V okviru tega projekta bodo pripravljeni trije ločeni sklopi učnih gradiv in gradiv za krepitev zmogljivosti.

(a) *Priročnik (1)* – V priročniku bodo zbrani osnovno znanje in sredstva za akterje na področju umetne inteligence, da v procesu raziskav in inovacij ocenijo in ublažijo tveganja preusmerjanja in zlorabe civilne tehnologije umetne inteligence s strani neodgovornih akterjev. Vseboval bo razprave o tem, zakaj in kako lahko odločitve v zvezi z razvojem in razširjanjem raziskav in inovacij vplivajo na tveganja preusmerjanja in zlorabe ter posledično ustvarjajo tveganja ali priložnosti za mir in varnost. Uvedel bo tudi ustrezne obveznosti po mednarodnem pravu in obveznosti nadzora izvoza, poleg vprašanj varnosti in zaščite, o katerih se razpravlja v vojaških krogih, pa tudi v krogih razoroževanja, nadzora nad oborožitvijo in neširjenja orožja; ter predstavil primere postopkov in orodij za izvajanje odgovornih inovacij, kot so metodologije za oceno učinka tehnologije in predloge za oceno tveganja.

(b) *Serija poddaj (~10)* – Te poddaje bodo delovale kot dostopen in privlačen medij za akterje na področju umetne inteligence, da bi izvedeli, zakaj in kako lahko odgovorni inovacijski procesi na področju umetne inteligence podpirajo mednarodni mir in varnost z zmanjševanjem tveganj, ki jih predstavljata preusmerjanje in zloraba s strani neodgovornih akterjev. V seriji bodo pregledane pomembne teme (npr. vzorec preusmerjanja in zlorabe raziskav in inovacij z dvojno/splošno rabo; humanitarni, strateški in politični izzivi, povezani z morebitno zlorabo civilnih raziskovalnih inovacij na področju umetne inteligence; izzivi, s katerimi se soočajo krogi razoroževanja, nadzora nad oboroževanjem in neširjenja orožja pri prizadevanjih za zmanjšanje tveganja; kako razvijati odgovorne inovacije na podlagi ocene tveganja; skladnost z nadzorom izvoza; zasnova zmanjšanja tveganja; odgovorno objavljanje; poznavanje strank; ter izkušnje s simulacijskimi vajami) in bodo strukturirane na podlagi razgovorov, ki jih bo projektna skupina opravila s predstavniki ustreznih skupnosti.

(c) *Serija blogov (9–10)* – Skupina bo pripravila urejeno serijo blogov, katere namen je povečati prepoznavnost prizadevanj, ki poskušajo premostiti meje med civilno usmerjeno „odgovorno umetno inteligenco“ ter skupnostma za nadzor nad oborožitvijo in neširjenje orožja. Serija blogov bo zagotovila platformo za razširjanje spoznanj, zamisli in rešitev v zvezi z odkrivanjem in obravnavanjem tveganj, povezanih s preusmerjanjem in zlorabo civilne umetne inteligence v procesu raziskav in inovacij. Namen bloga je predstaviti raznolikost razmišljanj in pogledov v sektorju umetne inteligence.

To gradivo se bo javno razširjalo prek spletišč izvajalskih akterjev, njihove prisotnosti v družbenih medijih ter prek neposredne komunikacije z ustreznimi akademskimi subjekti, civilnimi strokovnimi združenji za umetno inteligenco in drugimi ustreznimi skupinami.

3.1.2. Pričakovani rezultati projekta

Pričakuje se, da bo s tem projektom vzpostavljen nov sklop gradiv, s katerimi se lahko civilni strokovnjaki na področju umetne inteligence seznanijo s tem, a) kako se lahko civilne raziskave in inovacije na področju umetne inteligence preusmerijo in zlorabijo na načine, ki bi lahko pomenili tveganja za mednarodni mir in varnost, b) kako se taka tveganja obravnavajo v

skupnosti za razoroževanje, nadzor nad oborožitvijo in neširjenje orožja ter c) kako bi lahko strokovnjaki na področju umetne inteligence z odgovornimi inovacijskimi procesi dodatno prispevali k ublažitvi takih tveganj.

To bo predvidoma pospešilo večje sodelovanje civilnega sektorja na področju umetne inteligence pri zmanjševanju tveganj, ki jih lahko za mednarodni mir in varnost pomenita preusmerjanje in zloraba civilne umetne inteligence; izboljšalo zmogljivost tehničnih strokovnjakov za sodelovanje v ustreznih procesih na področju razoroževanja, nadzora nad oborožitvijo in neširjenja orožja ter podpiralo sodelovanje novih ciljnih skupin, ki običajno niso vključene v prizadevanja za izobraževanje o razoroževanju in neširjenju orožja.

Gradivo naj bi podpiralo tudi izvajanje drugih projektov in bo podlaga za izobraževalne dejavnosti in dejavnosti krepitve zmogljivosti v okviru projekta 2 ter dejavnosti dialoga in sodelovanja v okviru projekta 3. Pričakuje se, da bodo te dejavnosti prispevale k pripravi in izboljšanju gradiva. Tak iterativni pristop bo predvidoma pomagal odpraviti morebitne ovire za njihovo promocijo, razširjanje in uporabo v skupnosti umetne inteligence, vključno z vprašanji, povezanimi z jezikom, vsebino, kontekstom in razpoložljivostjo, ki bi lahko preprečile njihov vpliv na svetovni ravni, zlasti na svetovnem jugu.

3.2. Projekt št. 2 – izobraževalne dejavnosti in dejavnosti krepitve zmogljivosti za prihodnje generacije strokovnjakov na področju umetne inteligence

3.2.1. Namen projekta

Namen projekta št. 2 je podpirati vključevanje problematike preusmerjanja in zlorabe civilnih raziskav na področju umetne inteligence s strani neodgovornih akterjev v izobraževanje prihodnjih generacij strokovnjakov s tega področja. Dolgoročno bo to zagotovilo, da se bodo študentje naravoslovja, tehnologije, inženirstva in matematike, ki oblikujejo prihodnost umetne inteligence, zavedali negativnih učinkov na mednarodni mir in varnost, ki bi lahko nastali, če bi neodgovorni akterji preusmerjali in zlorabljali njihovo delo, ter da bodo imeli na razpolago osnovna orodja, potrebna za prepoznavanje in blaženje tovrstnih tveganj v procesu raziskav in inovacij.

V sodelovanju z izbranimi mednarodnimi univerzami in industrijskimi akterji se bo v okviru projekta z izobraževalci in študenti izvedla vrsta izobraževalnih delavnic in delavnic krepitve zmogljivosti. Namen projekta je tako razviti dejavnosti krepitve zmogljivosti, ki bi jih izobraževalci in oblikovalci akademskih učnih načrtov lahko vključili v usposabljanje prihodnjih generacij strokovnjakov s področja umetne inteligence (npr. tečaji o etiki umetne inteligence in odgovornih inovacijah) in v razmislek glede tveganj preusmerjanja in zlorabe civilnih raziskav in inovacij na področju umetne inteligence s strani neodgovornih akterjev ter jih povezali s širšim mirovnim in varnostnim kontekstom. S temi delavnicami bo projekt skušal opredeliti tudi mrežo zainteresiranih izobraževalcev, oblikovalcev učnih načrtov in študentov, ki bi spodbujali razširjanje in promocijo projektnega učnega gradiva in dejavnosti krepitve zmogljivosti v izobraževalni skupnosti na področju umetne inteligence, ter skupnosti strokovnjakov za umetno inteligenco. Namen te prvine mreženja je zagotoviti trajnost projektov tudi po njihovem zaključku in omogočiti vzpostavljanje močnejših povezav v podporo civilnim tehničnim

prizadevanjem za doseganje širših ciljev, ki zadevajo mir, varnost, razoroževanje in nadzor nad oborožitvijo.

3.2.2. Opis projekta

V okviru projekta se bo z izobraževalci in študenti z univerz širom po svetu izvedla vrsta izobraževalnih delavnic in delavnic krepitve zmogljivosti. Delavnice bi vključevale preplet predavanj in interaktivnih dejavnosti, v katerih bi izobraževalci in študentje imeli priložnosti za razmislek o tem, kako bi se lahko civilne raziskave in inovacije na področju umetne inteligence preusmerile in zlorabile na načine, ki bi ogrožali mednarodni mir in varnost, ter kako bi bilo mogoče tovrstna tveganja prepoznati, preprečiti ali ublažiti v procesu raziskav in inovacij ali z drugimi procesi upravljanja. Te dejavnosti bodo temeljile na pilotnem delu, ki ga je v manjšem obsegu predhodno izvedel Urad ZN za razorožitev (UN-ODA), ki je preskušal tako različne metode sodelovanja s študenti naravoslovja, tehnologije, inženirstva in matematike ter ozaveščanja teh študentov o tem, kako pomembno je razmišljati o širših učinkih njihovega dela, kot metode sodelovanja s strokovnjaki zunaj področja njihove stroke. Natančneje, delavnice bi vključevale:

(a) *regionalne delavnice krepitve zmogljivosti za izobraževalce in študente (4)* – z regionalnimi delavnicami se bodo izvajale in spodbujale dejavnosti, ki jih lahko izobraževalci uporabijo za krepitev zmogljivosti študentov naravoslovja, tehnologije, inženirstva in matematike v zvezi z odgovornimi inovacijami na področju umetne inteligence, s posebnim poudarkom na tem, kako oceniti in ublažiti tveganja, povezana s preusmerjanjem in zlorabo civilne tehnologije umetne inteligence s strani neodgovornih akterjev. Vsaka delavnica bo organizirana v sodelovanju z eno od univerz s sedežem v EU in eno od uglednih univerz iz druge svetovne regije, tako da bodo vedno združevale različne udeležence iz EU in udeležence iz držav zunaj EU. Delavnice bodo vključevale Latinsko Ameriko in Karibe, Severno Ameriko, Afriko, Azijo in Pacifik. Tako se bodo lahko delavnic udeležili tudi študentje (magistrskih in doktorskih študijev) z vseh koncev sveta, tudi s svetovnega juga. Delavnica bi potekala večinoma v angleščini, vendar bi imeli udeleženci, kjer bi bilo to izvedljivo, možnost, da sodelujejo v dejavnostih na osnovi drugih jezikovnih skupin; in

(b) *mednarodno delavnico o trajnostni krepitvi zmogljivosti (1)* – delavnica bo temeljila na izkušnjah, pridobljenih na regionalnih delavnicah, ter omogočila izmenjavo informacij in izkušenj med izobraževalci in izbranimi študenti z univerz, vključenih v projekt. Delavnica bo namenjena razpravi o tem, kako bi bilo mogoče dejavnosti in orodja, razvita med projektom, izpopolniti in razširiti zunaj okvirov skupine sodelujočih univerz. Razprava bo potekala tudi o tem, kako podpreti študente, da se bodo po vstopu na trg dela zavzemali za odgovorno delovanje na področju umetne inteligence, ki obravnava tveganja za mednarodni mir in varnost zaradi preusmerjanja in zlorabe.

Mreže in prisotnost SIPRI in UN-ODA v Afriki, azijsko-pacifiški regiji, Evropi, Severni in Južni Ameriki se bodo po potrebi uporabljale za omogočanje in podpiranje vidikov dejavnosti.

3.2.3. Pričakovani rezultati projekta

Projekt naj bi ustvaril modele dejavnosti krepitve zmogljivosti in dejavnosti sodelovanja, ki bi jih lahko izobraževalci in oblikovalci akademskih učnih načrtov posnemali, da bi prihodnje generacije strokovnjakov na področju umetne inteligence ozaveščali o problematiki preusmerjanja in zlorabe civilne umetne inteligence s strani neodgovornih akterjev ter o tem, kako lahko prispevajo k reševanju te problematike z odgovornimi inovacijskimi procesi. Pričakuje se, da bodo lahko udeleženci (izobraževalci in tudi študentje naravoslovja, tehnologije, inženirstva in matematike) po zaključku projektnih dejavnosti uporabljali in promovirali odgovorna inovacijska orodja, metode in koncepte za prepoznavanje in zmanjševanje tveganj preusmerjanja in zlorabe pri razvoju in razširjanju civilnih raziskav in inovacij na področju umetne inteligence.

Projektne dejavnosti naj bi prav tako ustvarile mrežo izobraževalcev, oblikovalcev učnih načrtov in študentov, ki ne bi le promovirali projektnih dejavnosti v izobraževalnih in strokovnih skupnostih na področju umetne inteligence (npr. na konferencah Inštituta inženirjev elektrotehnike in elektronike „Computational Intelligence Society“), temveč bi lahko s tehničnimi zmogljivostmi prispevali tudi k mednarodnim postopkom upravljanja, ki jih vodijo države (npr. postopek Konvencije o nekaterih vrstah klasičnega orožja v zvezi z nastajajočimi tehnologijami na področju smrtonosnih avtonomnih orožnih sistemov).

Kratkoročna in dolgoročna vrednost teh dejavnosti se bo pokazala z raziskavami pred izvedbo dejavnosti in po njej.

3.3. Projekt 3 – Spodbujanje dolgoročnejšega trajnostnega razvoja, razširjanja in učinkov odgovornih inovacij na področju umetne inteligence za mir in varnost

3.3.1. Namen projekta

Namen projekta 3 je spodbujati dolgoročnejši trajnostni razvoj, razširjanje in učinke odgovornih inovacij na področju umetne inteligence za zmanjšanje tveganj, ki jih lahko za mir in varnost predstavljata preusmerjanje in zloraba civilnih raziskav in inovacij na področju umetne inteligence. To se bo doseglo z okroglimi mizami s predstavniki industrije umetne inteligence, dialogi med več deležniki, pripravo javnega poročila in ciljno usmerjenimi dejavnostmi razširjanja. Cilj projekta je zagotoviti, da opravljeno delo, zlasti izobraževalne dejavnosti, dejavnosti krepitve zmogljivosti in dejavnosti sodelovanja, doseže in vpliva na širšo skupnost na področju umetne inteligence na vseh ravneh (kar zajema študente, inženirje in druge strokovnjake na področju umetne inteligence) ter prek vseh geografskih, sektorskih in drugih meja. Za doseganje najširšega in najglobljega učinka je bistveno sodelovati s strokovnimi organizacijami na tem področju, kot je Inštitut inženirjev elektrotehnike in elektronike, in vzpostaviti večdimenzionalno sodelovanje med akademskimi krogi, industrijo in drugimi skupinami. Taka prizadevanja bodo zainteresiranim predstavnikom iz različnih skupnosti na področju umetne inteligence omogočila, da prevzamejo odgovornost za problematiko in prispevajo svoje poglede na to, kako bi bilo mogoče uresničiti prizadevanja za zmanjšanje tveganja in ta prizadevanja na trajnosten način spodbujati v svetovni skupnosti na področju umetne inteligence. Poleg tega je za dolgoročno vrednost projekta za države, medvladne organizacije in druge pomembno, da se lahko strokovnjaki s področja umetne inteligence učijo

od vladnih strokovnjakov, ki se ukvarjajo z zmanjševanjem tveganj v okviru razoroževanja, nadzora nad oborožitvijo in neširjenja orožja, ter z njimi sodelujejo. Za trajnostnost je ključno tudi zagotoviti ustrezno analiziranje, utrjevanje in razširjanje znanj, pridobljenih z dejavnostmi sodelovanja.

3.3.2. Opis projekta

Ta projekt vključuje naslednja ključna področja delovanja:

(a) *dialoge med več deležniki o „odgovornih inovacijah na področju umetne inteligence za mir in varnost“ (do 9)* – niz dialogov, ki bodo potekali v virtualni obliki, bi združil strokovnjake iz akademskih krogov, zasebnega sektorja, s področja raziskav in tradicionalnega nadzora nad oborožitvijo, ki prihajajo iz EU in držav zunaj EU, da bi razpravljali o

i. tehnoloških trendih, ki lahko povzročijo tveganja širjenja, preusmerjanja ali zlorabe, kar ogroža mednarodni mir in varnost;
ii. načinih, na katere bi lahko zmanjšali tveganje z odgovornimi inovacijskimi procesi, metodami in sredstvi ter priložnostmi in izzivi za dialog in izmenjavo znanja med skupnostmi deležnikov, vključno s tistimi, ki delujejo v drugih sektorjih, kot sta biološka in kemična industrija; in
iii. potencialni vrednosti, namenu in formatu samostojne mreže strokovnjakov in dejavnosti dialoga. Skupina strokovnjakov se bo sestala večkrat letno in si prizadevala za organizacijo dveh javnih dogodkov za širšo skupnost.

Od devetih virtualnih srečanj naj bi bili dve odprti za javnost, da bi se omogočilo širše posvetovanje.

(b) *okrogle mize zasebnega sektorja (do 6)* – z nizom okroglih miz, ki bodo potekale v virtualni obliki, se bo začel dialog z akterji, ki delujejo na področju odgovornih inovacijskih postopkov na področju umetne inteligence v zasebnem sektorju (npr. partnerstvo za umetno inteligenco), in sicer o tem, kako lahko pomagajo zmanjšati tveganja preusmerjanja in zlorabe civilnih tehnologij umetne inteligence s strani neodgovornih akterjev, poleg tega pa se bodo z okroglimi mizami preučevale možnosti spodbujanja tovrstnih prizadevanj v okviru razvoja zasebnega sektorja. Teme bodo vključevale:

i. pomen konteksta mednarodne varnosti in razorožitve za zasebni sektor;
ii. pravno(-a) okolje(-a), v katerem(-ih) po vsem svetu potekajo razvoj, uvajanje in delovanje umetne inteligence;
iii. kako graditi na mehanizmih za oceno tveganja in drugih ukrepih, ki so del odgovornih inovacijskih postopkov in podjetniških programov za zagotavljanje skladnosti, ali pa bi lahko bili vanje vključeni; in
iv. izkušnje, pridobljene v drugih industrijah, postopkih in okvirih, ki se nanašajo na nadzor nad oborožitvijo (npr. biološka in kemična industrija).

(c) *poročilo o vidikih skupnosti na področju umetne inteligence glede nadzora nad oborožitvijo in zmanjševanja tveganja na področju umetne inteligence, ki je namenjeno skupnostim na področju umetne inteligence in nadzora nad oborožitvijo (1)* – priporočilo bo pripravljeno tako, da bodo ugotovitve in priporočila projekta zajeti in združeni v enoten referenčni dokument, namenjen civilno odgovorni umetni inteligenci in skupnostim na področju nadzora nad oborožitvijo. V poročilu bi bili opisani načini opredelitve, ocenjevanja in obravnave mednarodnih tveganja za mir in varnost, povezana s preusmerjanjem in zlorabo civilnih raziskav in inovacij na področju umetne inteligence.

(d) *dogodke za razširjanje informacij, namenjene posvetovanju in sodelovanju s skupnostmi na področju umetne inteligence in nadzora nad oborožitvijo (še ni določeno)* – skupina bo med projektom vseskozi iskala priložnosti za obveščanje o delu ter ugotovitvah in priporočilih, ki bodo iz tega izhajali. Format dogodkov in vsebina predstavitev bi bila prilagojena potrebam ciljnih skupin. To lahko vključuje srečanja Delovne skupine za neširjenje orožja in izvoz orožja (CONOP), skupščine evropskega zavezništva za umetno inteligenco, skupine vladnih strokovnjakov za nastajajoče tehnologije na področju smrtonosnih avtonomnih orožnih sistemov, medagencijske delovne skupine za umetno inteligenco (IAWG-AI), pobude Mednarodne telekomunikacijske zveze (ITU) „AI for Good“ (Umetna inteligenca za dobro), letnega dialoga o inovacijah Inštituta Združenih narodov za raziskave na področju razoroževanja (UNIDIR) in Inštituta inženirjev elektrotehnike in elektronike. Skupina bi si prizadevala tudi za dvostransko sodelovanje z ustreznimi deležniki iz vladnega sektorja, akademskih krogov in zasebnega sektorja.

3.3.3. Pričakovani rezultati projekta

Projekt naj bi postavil temelje za trajnostni razvoj, razširjanje in vpliv odgovornih inovacij v postopkih umetne inteligence, ki obravnavajo tveganja širjenja, preusmerjanja in zlorabe ter njihove posledice za mir in varnost tudi po neposrednem obdobju veljavnosti sklepa Sveta.

Dialog med več deležniki naj bi vzpostavil model za izmenjavo informacij in sodelovanje pri zmanjševanju tveganja ne le v svetovni skupnosti na področju umetne inteligence, temveč tudi med skupnostjo na področju civilno odgovorne umetne inteligence ter skupnostmi na področju razoroževanja, nadzora nad oborožitvijo in neširjenja orožja. Takšen model bi se lahko uporabljal za seznanjanje oblikovalcev politik s ključnimi tehnološkimi in znanstvenimi dosežki, pomembnimi za odgovorne inovacije na področju umetne inteligence, ter za seznanjanje občinstva iz tehničnih krogov z okoljem, v katerem trenutno sodelujejo oblikovalci politik. Projekt naj bi tudi omogočil trajnostne odnose in sodelovanje med zainteresiranimi akterji v teh različnih skupnostih in med njimi. Tovrstni heterogeni mrežni učinki naj bi omogočili večji razvoj in široko spodbujanje odgovornih inovacij na področju umetne inteligence za mir in varnost tudi zunaj časovnega okvira projekta.

Dialog zasebnega sektorja naj bi omogočil tesnejše in globlje sodelovanje zasebnega sektorja na področju umetne inteligence pri prepoznavanju, preprečevanju in zmanjševanju tveganj za mir in varnost, ki so posledica preusmerjanja in zlorabe civilnih raziskav in inovacij na področju umetne inteligence. Projekt naj bi ključnim akterjem v postopkih zasebnega sektorja omogočil

boljše razumevanje in prevzemanje odgovornosti za težave, ki jih želi rešiti. Poleg tega je njegov cilj omogočiti (širše) sprejetje in izvajanje odgovornih inovacijskih procesov, metod in sredstev v obstoječih mehanizmih in postopkih za upravljanje korporativnih tveganj.

Dialog med več deležniki in okrogla miza z zasebnim sektorjem naj bi prav tako omogočila vpogled v številna pomembna vprašanja, med drugim o tem, a) kako je mogoče nadalje izpopolniti in uporabiti odgovorne inovacijske metode in sredstva za prepoznavanje, preprečevanje in zmanjševanje tveganj zaradi preusmerjanja in zlorabe civilnih raziskav in inovacij na področju umetne inteligence, b) kako se lahko raziskave in inovacije na področju umetne inteligence na pozitiven način uporabijo v podporo mirovnim in varnostnim ciljem (npr. z aplikacijami za zgodnje opozarjanje na konflikte in humanitarno pomočjo), ter c) kako omogočiti širši dialog in izmenjavo informacij med prizadevanji za zmanjšanje tveganja v skupnosti na področju civilno odgovorne umetne inteligence (tj. različne pobude pod vodstvom Inštituta inženirjev elektrotehnike in elektronike) ter prizadevanji, ki se že izvajajo v skupnosti na področju razoroževanja, nadzora nad oborožitvijo in neširjenja orožja na medvladni ravni.

V okviru dejavnosti poročanja in razširjanja se bo analiziralo, utrdilo in razširjalo znanje, pridobljeno s projekti 1, 2 in 3, in tako bo podprta promocija ugotovitev in priporočil projektnih dejavnosti v svetovni skupnosti na področju umetne inteligence in na področju politik. Dejavnosti naj bi prav tako pomagale ohraniti učinek projektov tudi zunaj njihovega časovnega okvira.

4. TRAJANJE

Izvajanje projektov naj bi v celoti trajalo 36 mesecev.