



Consiglio
dell'Unione europea

Bruxelles, 11 novembre 2022
(OR. en)

14096/22
ADD 1

LIMITE

CORLX 996
CFSP/PESC 1433
CONUN 255
CODUN 49
CONOP 113
COTER 257
COARM 218

NOTA

Oggetto: DECISIONE DEL CONSIGLIO a sostegno dell'attuazione di un progetto di promozione dell'innovazione responsabile nell'intelligenza artificiale per la pace e la sicurezza - Allegato

ALLEGATO

DOCUMENTO DI PROGETTO

Promozione dell'innovazione responsabile nell'IA per la pace e la sicurezza

1. CONTESTO

I recenti progressi in materia di intelligenza artificiale (IA) hanno sbloccato nuove possibilità per sostenere e promuovere la pace e la sicurezza, ad esempio attraverso miglioramenti tecnologici in settori quali l'allerta rapida di prevenzione dei conflitti e la verifica del controllo degli armamenti e delle esportazioni. D'altro canto, questi progressi hanno aperto la strada a nuovi mezzi per generare o aggravare tensioni, conflitti e insicurezza tra gli Stati e al loro interno. I rischi posti da determinate applicazioni dell'IA, come i sistemi di armi letali autonomi, si sono rivelati una delle principali preoccupazioni per la comunità del controllo degli armamenti. Un rischio che merita maggiore attenzione e nei confronti del quale gli attuali sforzi diplomatici e di controllo degli armamenti potrebbero essere insufficienti è rappresentato dalla diversione e dall'uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili, compresi attori non statali malintenzionati, come l'uso improprio di reti generative avversarie (GAN) per produrre "deepfake" per le campagne di disinformazione.

L'IA è una tecnologia abilitante con un grande potenziale di uso generale. La ricerca e l'innovazione in materia di IA sviluppate per applicazioni civili potrebbero essere accessibili e riutilizzate (in modo relativamente semplice) da determinati attori per usi dannosi o destabilizzanti, che potrebbero avere implicazioni per la pace e la sicurezza internazionali. La diversione e l'uso improprio delle tecnologie civili non sono fenomeni nuovi e non riguardano solo l'IA. Nel settore correlato della robotica un esempio recente è rappresentato dalla trasformazione in armi e dall'utilizzo dei droni ad uso ricreativo da parte dell'ISIS/Daesh. Tuttavia, nel caso dell'IA il problema è complicato da molteplici fattori: la natura immateriale e in rapido mutamento degli algoritmi e dei dati dell'IA, che rende difficile controllarne il trasferimento e la proliferazione; il ruolo guida del settore privato nell'ecosistema della ricerca, dello sviluppo e dell'innovazione e la conseguente necessità di proteggere gli algoritmi proprietari; la disponibilità globale delle competenze umane e delle risorse materiali in grado di riconvertire le tecnologie dell'IA. Intanto, troppo spesso chi lavora nell'IA nel settore civile non è consapevole delle potenziali implicazioni che la diversione e l'uso improprio del suo lavoro potrebbero avere per la pace e la sicurezza internazionali o è restio a partecipare alle discussioni sui rischi dell'IA in corso negli ambienti del controllo degli armamenti e della non proliferazione.

È necessario sostenere un maggiore impegno della comunità civile dell'IA nella comprensione e nell'attenuazione dei rischi per la pace e la sicurezza associati alla diversione e all'uso improprio della tecnologia civile dell'IA da parte di attori irresponsabili. Per l'Istituto internazionale di ricerca sulla pace di Stoccolma (SIPRI) e l'Ufficio per gli affari del disarmo delle Nazioni Unite (UNODA), tale obiettivo potrebbe essere conseguito attraverso una maggiore promozione dell'innovazione responsabile nella comunità civile mondiale dell'IA. I lavori svolti in passato dal SIPRI e dall'UNODA hanno dimostrato che un'innovazione responsabile, in quanto meccanismo di autogoverno, potrebbe fornire alla comunità civile mondiale dell'IA strumenti e metodi pratici per individuare e contribuire a prevenire e attenuare i rischi che la diversione e l'uso improprio della ricerca e dell'innovazione civili in materia di IA potrebbero comportare per la pace e la sicurezza. I lavori del SIPRI e dell'UNODA hanno inoltre individuato metodologie e diverse iniziative in corso incentrate sul settore civile in materia di IA responsabile che potrebbero essere sviluppate per sensibilizzare la comunità civile dell'IA riguardo alle questioni relative al controllo degli armamenti e alla non proliferazione, ai dibattiti di esperti e alle posizioni degli Stati in materia di sviluppo responsabile, diffusione e utilizzo dell'IA, nonché insegnamenti da trarre dal lavoro di responsabilità del settore della difesa¹. Significativamente, questi lavori precedenti hanno chiaramente individuato nella cooperazione con gli studenti di scienza, tecnologia, ingegneria e matematica (STEM), che continuano a impegnarsi nel campo dell'IA in un contesto educativo, un elemento centrale di qualsiasi sforzo efficace di innovazione responsabile.

¹ Le metodologie comprendono, ad esempio, le pratiche raccomandate dall'Istituto degli ingegneri elettronici ed elettrotecnici (IEEE) per valutare l'impatto dei sistemi autonomi e intelligenti sul benessere umano (IEEE Std 7010-2020) e l'elenco di valutazione per un'intelligenza artificiale affidabile (ALTAI) del gruppo di esperti ad alto livello sull'intelligenza artificiale. Le iniziative proposte comprendono: l'iniziativa globale dell'IEEE sull'etica dei sistemi autonomi e intelligenti, il partenariato sull'IA e il partenariato globale sull'intelligenza artificiale.

2. OBIETTIVI

Tali progetti mirano a sostenere un maggiore coinvolgimento della comunità civile dell'IA nell'attenuare i rischi che la diversione e l'uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili possono comportare per la pace e la sicurezza internazionali. A tal fine essi intendono, in primo luogo, generare una maggiore comprensione del modo in cui le decisioni relative allo sviluppo e alla diffusione della ricerca e dell'innovazione in materia di IA possono incidere sui rischi di diversione e uso improprio, generando a loro volta rischi o opportunità per la pace e la sicurezza e, in secondo luogo, promuovere processi, metodi e strumenti di innovazione responsabili che possano contribuire a garantire l'applicazione pacifica delle innovazioni civili e la diffusione responsabile delle conoscenze in materia di IA. A tale scopo, sostengono attività di sviluppo di capacità, ricerca e dialogo che *i*) rafforzeranno la capacità all'interno della comunità civile mondiale dell'IA di includere e affrontare i rischi per la pace e la sicurezza derivanti dalla diversione e dall'uso improprio dell'IA civile da parte di attori irresponsabili attraverso processi di innovazione responsabili e *ii*) potenzieranno il collegamento tra gli sforzi di attenuazione dei rischi nell'IA responsabile in ambito civile e quelli già in corso nella comunità del disarmo, del controllo degli armamenti e della non proliferazione a livello intergovernativo. È importante rilevare che non intendono stabilire nuove norme, principi o regolamenti, né interferire in altro modo in settori di competenza degli Stati. Intendono invece sviluppare sforzi civili in materia di innovazione responsabile per includere i rischi per la pace e la sicurezza derivanti dalla diversione e dall'uso improprio dell'IA civile da parte di attori irresponsabili e offrire opportunità formative sui pertinenti sforzi intergovernativi esistenti.

Per raggiungere la comunità civile dell'IA e avere un impatto efficace su di essa, i progetti adottano un triplice approccio che mira a:

- a) *dialogare con gli educatori* — lavorare con educatori e sviluppatori di curricula universitari selezionati allo sviluppo e alla promozione di materiale didattico che possa essere utilizzato per integrare nella formazione dei futuri operatori del settore dell'IA (ad esempio nei corsi sull'etica e sull'innovazione responsabile in materia di IA) la presa in considerazione dei rischi per la pace e la sicurezza derivanti dalla diversione e dall'uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili;

- b) *dialogare con gli studenti* — spiegare a studenti selezionati di scienza, tecnologia, ingegneria e matematica (STEM) di tutto il mondo come i rischi per la pace e la sicurezza derivanti dalla diversione e dall'uso improprio dello sviluppo civile dell'IA da parte di attori irresponsabili possono essere individuati, evitati o attenuati nel processo di ricerca e innovazione o attraverso altri processi di governance e
- c) *dialogare con il settore dell'IA* — lavorare con associazioni professionali e organismi di normazione come l'Istituto degli ingegneri elettronici ed elettrotecnici (IEEE) per i) diffondere tra i professionisti tecnici materiale didattico mirato e attività di coinvolgimento, ii) sostenere gli usi positivi dell'IA per la pace e la sicurezza e iii) facilitare il dialogo e la condivisione di informazioni tra esperti del mondo accademico, del settore privato e del governo su come attenuare il rischio di diversione e uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili.

Tale approccio consente ai progetti di raggiungere la comunità dell'IA a tutti i livelli, comprese le generazioni future e non solo gli operatori attuali. Consente inoltre il coinvolgimento trasversale del mondo accademico, dell'industria e di altri settori e sostiene la sostenibilità degli sforzi futuri creando reti che attraversano i confini tra loro.

I progetti mirano inoltre a sfruttare la capacità di mobilitazione e l'esperienza del SIPRI e dell'UNODA per incidere sulla comunità dell'IA a livello mondiale e non solo sui portatori di interessi dell'UE. Il SIPRI e l'UNODA si trovano in una posizione privilegiata per raggiungere gli attori dell'IA provenienti da tutta l'Africa, l'Asia-Pacifico, l'Europa e l'America settentrionale e meridionale e facilitare il dialogo tra loro. Entrambi gli enti hanno inoltre esperienza di lavoro in altri settori scientifici e tecnologici che si trovano ad affrontare sfide analoghe di duplice uso e proliferazione, compresa la biotecnologia. I progetti cercano inoltre di sfruttare le condizioni presenti nell'Unione europea, quali a) l'esistenza di processi multilaterali avanzati in materia di IA responsabile, b) l'elevato livello di coinvolgimento e di competenza in materia di disarmo, controllo degli armamenti e non proliferazione nell'UE, c) la diversità dei legami che le organizzazioni del mondo accademico, della ricerca e del settore privato dell'UE intrattengono con altre regioni, in particolare nel Sud globale, che costituiranno un ulteriore obiettivo importante di coinvolgimento e d) la varietà delle cittadinanze degli studenti, degli educatori e degli ingegneri nelle università, negli istituti di ricerca e nel settore privato.

L'inclusione sarà una considerazione fondamentale per lo svolgimento delle attività dei progetti. Per sostenere efficacemente la comunità dell'IA, i progetti riconoscono che essa è costituita da una gamma diversificata di attori, in particolare che il genere è un fattore di grande rilevanza.

- a) Per questo motivo, la dimensione di genere sarà integrata in linea con le strategie di integrazione della dimensione di genere e di parità di genere a livello del sistema delle Nazioni Unite. La partecipazione delle donne a tutte le attività del progetto sarà incoraggiata e richiesta e
- b) l'inclusione delle persone con disabilità e soluzioni ragionevoli per soddisfare le loro esigenze saranno assicurati durante tutte le fasi del progetto. Saranno pertanto eliminati gli ostacoli alla partecipazione delle persone con disabilità e si provvederà ad adottate misure volte a coinvolgerle e a facilitare la presentazione delle loro opinioni ed esperienze sostanziali.

3. PROGETTI

I tre progetti descritti di seguito sono destinati a essere complementari e a sostenersi reciprocamente, con elementi che coprono l'intero arco di 36 mesi.

3.1. Progetto 1 — Produzione di materiale didattico e di sviluppo di capacità per la comunità civile dell'IA

3.1.1. Finalità del progetto

L'obiettivo primario del progetto 1 è fornire le conoscenze e i mezzi affinché gli attori civili dell'IA possano valutare e attenuare i rischi che la diversione e l'uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili possono comportare per la pace e la sicurezza internazionali. Il progetto mira a produrre materiale didattico e di sviluppo di capacità che fornirà agli operatori dell'IA di tutte le regioni, di tutti i livelli e di tutti i settori (compresi gli educatori incentrati sull'IA, gli sviluppatori di curricula universitari, gli studenti nei settori STEM e gli ingegneri e i ricercatori in materia di IA nel mondo accademico e nel settore privato) le informazioni e gli strumenti necessari per

- a)* comprendere in che modo la ricerca e l'innovazione civili in materia di IA potrebbero essere oggetto di diversione e uso improprio in modi che potrebbero presentare rischi per la pace e la sicurezza internazionali e in che modo le decisioni relative allo sviluppo e alla diffusione della ricerca e dell'innovazione possono aumentare o ridurre il rischio di diversione e uso improprio;
- b)* comprendere gli sforzi già intrapresi dalla comunità del disarmo, del controllo degli armamenti e della non proliferazione per attenuare i rischi di diversione e uso improprio della ricerca e dell'innovazione civili e
- c)* praticare un'innovazione responsabile in modo da attenuare il rischio di diversione e uso improprio nello sviluppo e nella diffusione della ricerca e dell'innovazione.

3.1.2. Descrizione del progetto

Questo progetto produrrà tre serie distinte di materiale didattico e di sviluppo di capacità.

- a) *Manuale (1)* — Il manuale rappresenterà una raccolta delle conoscenze di base e dei mezzi che permetteranno agli attori dell'IA di valutare e attenuare, nel processo di ricerca e innovazione, i rischi di diversione e uso improprio della tecnologia civile dell'IA da parte di attori irresponsabili. Discuterà di come e perché le decisioni relative allo sviluppo e alla diffusione della ricerca e dell'innovazione possono incidere sui rischi di diversione e uso improprio e, a loro volta, generare rischi o opportunità per la pace e la sicurezza. Illustrerà inoltre gli obblighi pertinenti previsti dal diritto internazionale e dal controllo delle esportazioni, oltre a considerazioni in materia di sicurezza e protezione in corso di discussione negli ambienti militari e negli ambienti del disarmo, del controllo degli armamenti e della non proliferazione, e presenterà esempi di processi e strumenti per la pratica dell'innovazione responsabile, come metodologie di valutazione dell'impatto tecnologico e modelli di valutazione dei rischi.

- b) *Serie di podcast (~10)* — Questi podcast saranno un mezzo accessibile e coinvolgente grazie al quale gli attori dell'IA potranno capire come e perché i processi di innovazione responsabile in materia di IA possono sostenere la pace e la sicurezza internazionali attraverso l'attenuazione dei rischi derivanti dalla diversione e dall'uso improprio da parte di attori irresponsabili. La serie affronterà temi importanti (ad esempio il modello di diversione e uso improprio della ricerca e dell'innovazione a duplice uso/uso generale, le sfide umanitarie, strategiche e politiche associate al potenziale uso improprio dell'innovazione e della ricerca civili in materia di IA, le sfide che gli ambienti del disarmo, del controllo degli armamenti e della non proliferazione devono affrontare negli sforzi di attenuazione dei rischi, come realizzare un'innovazione responsabile attraverso la valutazione dei rischi, il rispetto delle norme in materia di controllo delle esportazioni, la riduzione del rischio fin dalla progettazione, la pubblicazione responsabile, la conoscenza dei clienti e l'esperienza acquisita tramite esercizi di simulazione) e sarà strutturata intorno a interviste che il gruppo incaricato del progetto condurrà con i rappresentanti delle comunità interessate.

- c) *Serie di blog post (9-10)* — Il team svilupperà una serie di blog post selezionati volta ad accrescere la visibilità degli sforzi intesi ad attraversare i confini tra le comunità di "IA responsabile" incentrate sul settore civile e le comunità del controllo degli armamenti e della non proliferazione. La serie di blog post fornirà una piattaforma per diffondere informazioni, idee e soluzioni riguardanti l'individuazione e la gestione dei rischi associati alla diversione e all'uso improprio dell'IA civile nel processo di ricerca e innovazione. Il blog cercherà di rappresentare la diversità di pensiero e di prospettive presente nel settore dell'IA.

Tale materiale sarà divulgato pubblicamente attraverso i siti web dei soggetti responsabili dell'attuazione, la loro presenza sui social media e la comunicazione diretta con gli enti accademici pertinenti, le associazioni professionali civili dell'IA e altri gruppi appropriati.

3.1.3. Risultati attesi del progetto

Il progetto dovrebbe stabilire una nuova serie di materiali attraverso i quali sensibilizzare gli operatori civili dell'IA a) su come la ricerca e l'innovazione civili in materia di IA potrebbero essere oggetto di diversione e uso improprio secondo modalità tali da presentare rischi per la pace e la sicurezza internazionali, b) su come tali rischi sono affrontati dalla comunità del disarmo, del controllo degli armamenti e della non proliferazione e c) su come gli operatori dell'IA potrebbero contribuire ulteriormente all'attenuazione di tali rischi attraverso processi di innovazione responsabili.

Ciò dovrebbe consentire di portare avanti il coinvolgimento del settore civile dell'IA nell'attenuare i rischi che la diversione e l'uso improprio dell'IA civile possono comportare per la pace e la sicurezza internazionali, migliorare la capacità degli operatori tecnici di impegnarsi nei processi pertinenti in materia di disarmo, controllo degli armamenti e non proliferazione e sostenere il coinvolgimento di nuovi pubblici non tradizionalmente inclusi negli sforzi di istruzione in materia di disarmo e non proliferazione.

Il materiale dovrebbe inoltre sostenere l'attuazione degli altri progetti e servirà da base per le attività in materia di istruzione e sviluppo di capacità nell'ambito del progetto 2, nonché per le attività di dialogo e coinvolgimento nell'ambito del progetto 3. Tali attività dovrebbero a loro volta alimentare la produzione e il perfezionamento del materiale. Questo approccio iterativo dovrebbe contribuire ad affrontare i potenziali ostacoli alla loro promozione, alla loro diffusione e al loro utilizzo all'interno della comunità dell'IA, tra cui le questioni connesse al linguaggio, al contenuto, al contesto e alla disponibilità, che ne comprometterebbero l'impatto a livello globale, in particolare nel Sud globale.

3.2. Progetto 2 — Attività in materia di istruzione e sviluppo di capacità per i futuri operatori del settore dell'IA

3.2.1. Finalità del progetto

L'obiettivo del progetto 2 è sostenere l'integrazione del problema della diversione e dell'uso improprio della ricerca civile in materia di IA da parte di attori irresponsabili nell'istruzione delle future generazioni di operatori del settore dell'IA. A lungo termine, si garantirà così che gli studenti di scienza, tecnologia, ingegneria e matematica (STEM), che determinano il futuro dell'IA, siano consapevoli degli effetti negativi che la diversione e l'uso improprio del loro lavoro da parte di attori irresponsabili potrebbero avere sulla pace e sulla sicurezza internazionali e che dispongano degli strumenti di base necessari per individuare e attenuare tale rischio nel processo di ricerca e innovazione.

Nel quadro del progetto verrà organizzata una serie di seminari in materia di istruzione e sviluppo di capacità con educatori e studenti in collaborazione con università e attori del settore internazionali selezionati. Il progetto mira pertanto a elaborare attività in materia di sviluppo di capacità che gli educatori e gli sviluppatori di curricula universitari potrebbero utilizzare per includere nella formazione dei futuri operatori del settore dell'IA (ad esempio corsi sull'etica e sull'innovazione responsabile in materia di IA) la presa in considerazione dei rischi di diversione e uso improprio della ricerca e dell'innovazione civili in materia di IA da parte di attori irresponsabili e collegarle al più ampio contesto di pace e sicurezza. Tramite i seminari, il progetto mirerà inoltre a individuare una rete di educatori, sviluppatori di curricula universitari e studenti interessati a sostenere la diffusione e la promozione del materiale didattico del progetto e delle attività di sviluppo di capacità nella comunità dell'istruzione in materia di IA e nella comunità di operatori del settore dell'IA. Questo elemento di creazione di una rete mira a garantire la sostenibilità dei progetti al di là della loro durata immediata e a consentire la creazione di legami più forti a sostegno di un impegno tecnico civile a favore di obiettivi più ampi in materia di pace, sicurezza, disarmo e controllo degli armamenti.

3.2.2. Descrizione del progetto

Nel quadro del progetto verrà organizzata una serie di seminari in materia di istruzione e sviluppo di capacità con educatori e studenti di università selezionate di tutto il mondo. I seminari consisteranno in una combinazione di conferenze e attività interattive che offriranno agli educatori e agli studenti l'opportunità di riflettere sui modi in cui la ricerca e l'innovazione civili in materia di IA potrebbero essere oggetto di diversione e uso improprio in forme che potrebbero presentare rischi per la pace e la sicurezza internazionali e sui modi per individuare, prevenire o attenuare tali rischi nel processo di ricerca e innovazione o attraverso altri processi di governance. Tali attività si baseranno su precedenti attività pilota su scala ridotta svolte dall'UNODA, in cui si sono sperimentati metodi per dialogare con gli studenti STEM e sensibilizzarli all'importanza di tenere in considerazione l'impatto più ampio del loro lavoro e di avvalersi di competenze al di fuori del loro settore di specializzazione. In concreto, si tratterà di:

- a) seminari regionali in materia di sviluppo di capacità per educatori e studenti (4) — nell'ambito dei seminari regionali si svolgeranno e promuoveranno attività che gli educatori possano utilizzare per sviluppare le capacità degli studenti nei settori STEM in materia di innovazione responsabile dell'IA, con particolare attenzione alle modalità per valutare e attenuare i rischi di diversione e uso improprio della tecnologia civile in materia di IA da parte di attori irresponsabili. Ciascun seminario sarà organizzato con un'università con sede nell'UE e un'università di alto profilo di una diversa regione del mondo; in tal modo si collegheranno sempre un gruppo diversificato di partecipanti residenti nell'UE e un gruppo diversificato di partecipanti residenti al di fuori dell'UE. I seminari coinvolgeranno America latina e Caraibi, America settentrionale, Africa e Asia-Pacifico. In tal modo potranno partecipare studenti (a livello di master e dottorato) di tutto il mondo, compreso il Sud globale. I seminari si svolgeranno principalmente in inglese, ma, ove possibile, i partecipanti avranno la possibilità di partecipare ad attività sulla base di altri raggruppamenti linguistici; e
- b) seminario internazionale sullo sviluppo sostenibile di capacità (1) — il seminario si baserà sugli insegnamenti tratti dai seminari regionali e faciliterà lo scambio di informazioni ed esperienze tra educatori e studenti selezionati delle università coinvolte nel progetto. Nell'ambito del seminario si esamineranno i modi per perfezionare le attività e gli strumenti elaborati nel corso del progetto e diffonderli al di fuori del gruppo di università partecipanti. Si discuteranno inoltre i modi per sostenere il coinvolgimento degli studenti, una volta entrati nel mondo del lavoro, in un'IA responsabile che affronti i rischi di diversione e uso improprio per la pace e la sicurezza internazionali.

Le reti del SIPRI e dell'UNODA e la loro presenza in Africa, Asia-Pacifico, Europa, America settentrionale e meridionale saranno utilizzate per facilitare e sostenere, se del caso, determinati aspetti delle attività.

3.2.3. Risultati attesi del progetto

Il progetto dovrebbe creare modelli di attività di sviluppo di capacità e di coinvolgimento che gli educatori e gli sviluppatori di curricula universitari potrebbero replicare per sensibilizzare i futuri operatori del settore dell'IA ai problemi della diversione e dell'uso improprio dell'IA civile da parte di attori irresponsabili e ai modi in cui possono contribuire ad attenuare tali problemi attraverso processi di innovazione responsabile. Dopo il completamento delle attività del progetto, i partecipanti (educatori ma anche studenti nei settori STEM) dovrebbero essere in grado di utilizzare e promuovere strumenti, metodi e concetti di innovazione responsabile per individuare e attenuare i rischi di diversione e uso improprio nello sviluppo e nella diffusione della ricerca e dell'innovazione civili in materia di IA.

Si prevede inoltre che le attività del progetto portino alla formazione di una rete di educatori, sviluppatori di curricula universitari e studenti che non solo promuoverebbero le attività del progetto nelle comunità educative e professionali in materia di IA (ad esempio durante le conferenze dell'IEEE Computational Intelligence Society), ma potrebbero anche contribuire con capacità tecniche ai processi di governance internazionale a guida statale (ad esempio i processi della convenzione su certe armi convenzionali sulle tecnologie emergenti nel settore dei sistemi di armi letali autonomi).

Il valore a breve e lungo termine di tali attività sarà dimostrato mediante indagini pre- e post-attività.

3.3. Progetto 3 — Facilitare lo sviluppo sostenibile, la diffusione e l'impatto a lungo termine dell'innovazione responsabile nel settore dell'IA per la pace e la sicurezza

3.3.1. Finalità del progetto

L'obiettivo del progetto 3 è facilitare lo sviluppo sostenibile, la diffusione e l'impatto a lungo termine dell'innovazione responsabile nel settore dell'IA quale mezzo per attenuare i rischi che la diversione e l'uso improprio della ricerca e dell'innovazione civili in materia di IA possono comportare per la pace e la sicurezza. Il progetto intende ottenere tale risultato mediante tavole rotonde con il settore dell'IA, dialoghi multilaterali, l'elaborazione di una relazione pubblica e attività di diffusione mirate. Mira a garantire che il lavoro generato, in particolare le attività in materia di istruzione, sviluppo di capacità e coinvolgimento, raggiunga la comunità dell'IA in generale e incida su di essa, a tutti i livelli (dagli studenti agli ingegneri e altri professionisti dell'IA) e al di là dei confini geografici, settoriali e di altro tipo. Per aumentare la possibilità di produrre un impatto ampio e profondo, è essenziale cooperare con le organizzazioni professionali in questo settore, come l'IEEE, e condurre dialoghi multidimensionali e trasversali nel mondo accademico, nell'industria e in altri settori. Tali sforzi offriranno ai rappresentanti interessati delle diverse comunità dell'IA l'opportunità di assumere la titolarità del problema e di esprimere il proprio punto di vista sul modo in cui gli sforzi di attenuazione dei rischi possono essere realizzati e promossi in modo sostenibile sia all'interno della comunità mondiale dell'IA che tra le sue varie componenti. È inoltre importante per il valore a lungo termine del progetto per gli Stati, le organizzazioni intergovernative e altri soggetti, che gli operatori del settore dell'IA possano imparare da esperti governativi impegnati nell'attenuazione dei rischi nel contesto del disarmo, del controllo degli armamenti e della non proliferazione e dialogare con loro. È inoltre fondamentale per la sostenibilità garantire che le conoscenze generate dalle attività di coinvolgimento siano analizzate, consolidate e diffuse in modo adeguato.

3.3.2. Descrizione del progetto

Il progetto si articola in questi filoni principali:

- a) *dialoghi multilaterali su "Innovazione responsabile dell'IA per la pace e la sicurezza" (fino a 9)* — questa serie di riunioni virtuali di dialogo riunirà esperti del mondo accademico, della ricerca, del settore privato ed esperti del controllo tradizionale degli armamenti provenienti dall'UE e da paesi terzi perché discutano dei seguenti elementi:
- i. le tendenze a livello tecnologico che possono generare rischi di diffusione, diversione e uso improprio con ripercussioni sulla pace e la sicurezza internazionali;
 - ii. i modi per operare nell'attenuazione dei rischi attraverso processi, metodi e mezzi di innovazione responsabile nonché le opportunità e le sfide per il dialogo e la condivisione delle conoscenze tra le comunità dei portatori di interessi, fra cui quelli che operano in altri settori quali quello biologico e quello chimico; e
 - iii. il valore potenziale, lo scopo e il formato di una rete autosufficiente di esperti e attività di dialogo. Il gruppo di esperti si riunirà varie volte all'anno e lavorerà per organizzare due eventi pubblici per la comunità in generale.

Delle nove riunioni virtuali, due dovrebbero essere aperte al pubblico, al fine di facilitare una consultazione più ampia.

- b) *tavole rotonde del settore privato (fino a 6)* — questa serie di tavole rotonde virtuali avvierà un dialogo con gli attori che lavorano su processi di innovazione responsabile dell'IA nel settore privato (ad esempio il partenariato sull'IA) in merito al modo in cui possono contribuire a ridurre al minimo i rischi di diversione e uso improprio delle tecnologie civili dell'IA da parte di attori irresponsabili, e studierà possibili incentivi nello sviluppo del settore privato a tal fine. I temi comprenderanno:
- i. l'importanza del contesto internazionale in materia di sicurezza e disarmo per il settore privato;
 - ii. il contesto o i contesti giuridici in tutto il mondo nei quali ci si occupa di sviluppo, diffusione e funzionamento dell'IA;

- iii. come basarsi sui meccanismi di valutazione dei rischi e su altre misure che rientrano o potrebbero essere integrate nei processi di innovazione responsabile e nei programmi di conformità aziendale; e
- iv. gli insegnamenti da trarre da altri settori, processi e quadri relativi al controllo degli armamenti (ad esempio i settori biologico e chimico).
- c) relazione sulle prospettive della comunità dell'IA in materia di controllo degli armamenti e attenuazione dei rischi nell'IA, rivolta alle comunità dell'IA e del controllo degli armamenti (1) — nell'elaborazione di tale relazione si raccoglieranno e consolideranno i risultati e le raccomandazioni del progetto in un unico documento di riferimento destinato sia alla comunità civile dell'IA responsabile che alla comunità del controllo degli armamenti. Nella relazione si esamineranno i modi per individuare, valutare e affrontare i rischi per la pace e la sicurezza internazionali associati alla diversione e all'uso improprio della ricerca e dell'innovazione civili in materia di IA.
- d) eventi di diffusione mirati alla consultazione e al dialogo con le comunità dell'IA e del controllo degli armamenti (da stabilire) — il gruppo studierà occasioni di comunicare il lavoro e i relativi risultati nonché formulare raccomandazioni nel corso dell'intero progetto. Il formato degli eventi e il contenuto delle presentazioni saranno adattati alle esigenze dei gruppi destinatari e possono comprendere riunioni del CONOP, dell'Assemblea dell'Alleanza europea per l'IA, del gruppo di esperti governativi sulle tecnologie emergenti nel settore dei sistemi d'arma autonomi letali, del gruppo di lavoro interagenzie sull'IA (IAWG-AI), dell'iniziativa dell'UIT "AI for Good", del dialogo annuale sull'innovazione dell'UNIDIR e dell'Istituto degli ingegneri elettronici ed elettrotecnici. Il gruppo si adopererà inoltre per dialogare a livello bilaterale con i pertinenti portatori di interessi dei governi, del mondo accademico e del settore privato.

3.3.3. Risultati attesi del progetto

Il progetto dovrebbe gettare le basi per lo sviluppo sostenibile, la diffusione e l'impatto dell'innovazione responsabile dei processi di IA, affrontando i rischi di diffusione, diversione e uso improprio e le loro implicazioni per la pace e la sicurezza al di là della durata immediata della decisione del Consiglio.

Il dialogo multilaterale dovrebbe fornire un modello per la condivisione delle informazioni e la collaborazione in materia di attenuazione dei rischi non solo all'interno della comunità mondiale dell'IA, ma anche tra la comunità civile dell'IA responsabile e le comunità del disarmo, del controllo degli armamenti e della non proliferazione. Tale modello potrebbe essere utilizzato per familiarizzare i responsabili politici con i principali progressi tecnologici e scientifici pertinenti per l'innovazione responsabile dell'IA e anche per familiarizzare i tecnici con il contesto in cui attualmente operano i responsabili politici. Il progetto dovrebbe facilitare le relazioni e il dialogo sostenibili tra gli attori interessati all'interno di queste diverse comunità e tra di esse. Tali effetti di rete eterogenei dovrebbero consentire un maggiore sviluppo e un'ampia promozione dell'innovazione responsabile dell'IA per la pace e la sicurezza oltre l'orizzonte temporale del progetto.

Il dialogo con il settore privato dovrebbe consentire un coinvolgimento maggiore e più profondo del settore privato dell'IA nell'individuazione, nella prevenzione e nell'attenuazione dei rischi per la pace e la sicurezza derivanti dalla diversione e dall'uso improprio della ricerca e dell'innovazione civili in materia di IA. Il progetto dovrebbe offrire ai principali attori dei processi nel settore privato una maggiore comprensione e titolarità dei problemi che intende affrontare. Inoltre, mira a facilitare l'adozione e l'attuazione (più ampie) di processi, metodi e mezzi di innovazione responsabile nei meccanismi e nelle procedure aziendali di gestione del rischio esistenti.

Il dialogo multilaterale e la tavola rotonda del settore privato dovrebbero inoltre generare conoscenze su una serie di questioni sostanziali, tra cui a) il modo in cui metodi e mezzi di innovazione responsabile possono essere ulteriormente perfezionati e utilizzati per individuare, prevenire e attenuare i rischi derivanti dalla diversione e dall'uso improprio della ricerca e dell'innovazione civili in materia di IA, b) il modo in cui la ricerca e l'innovazione in materia di IA possono essere utilizzate in modo proficuo a sostegno degli obiettivi di pace e sicurezza (ad esempio, applicazioni per l'allerta rapida di prevenzione dei conflitti e l'assistenza umanitaria) e c) i modi per facilitare un maggiore dialogo e una maggiore condivisione di informazioni fra i vari sforzi di attenuazione dei rischi intrapresi nella comunità civile dell'IA responsabile (ossia varie iniziative a guida dell'IEEE) e quelli già in corso nella comunità del disarmo, del controllo degli armamenti e della non proliferazione a livello intergovernativo.

La relazione e le attività di diffusione analizzeranno, consolideranno e diffonderanno le conoscenze generate dai progetti 1, 2 e 3, sostenendo in tal modo la promozione dei risultati e delle raccomandazioni risultanti dalle attività del progetto all'interno della comunità mondiale dell'IA e nell'ambito della comunità politica. Dovrebbero inoltre contribuire a garantire la sostenibilità dell'impatto oltre la durata dei progetti.

4. DURATA

La durata totale stimata di attuazione dei progetti è di 36 mesi.
