



Rada
Unii Europejskiej

Bruksela, 26 października 2022 r.
(OR. en)

14095/22
ADD 1

LIMITE

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

WNIOSEK

Od:	Wysoki przedstawiciel Unii do spraw zagranicznych i polityki bezpieczeństwa (podpisał sekretarz generalny Stefano SANNINO)
Data otrzymania:	26 października 2022 r.
Do:	Thérèse BLANCHET, sekretarz generalna Rady Unii Europejskiej
Dotyczy:	Wniosek Wysokiego Przedstawiciela Unii do Spraw Zagranicznych i Polityki Bezpieczeństwa skierowany do Rady dotyczący decyzji Rady wspierającej realizację projektu pt. „Propagowanie odpowiedzialnych innowacji w zakresie sztucznej inteligencji na rzecz pokoju i bezpieczeństwa”

Delegacje otrzymują w załączeniu dokument HR(2022) 238.

Załącznik: HR(2022) 238

EUROPEJSKA SŁUŻBA DZIAŁAŃ ZEWNĘTRZNYCH



DOKUMENT PROJEKTOWY

Propagowanie odpowiedzialnych innowacji w zakresie sztucznej inteligencji na rzecz pokoju i bezpieczeństwa

HR(2022)238

1. KONTEKST

Ostatnie postępy w dziedzinie sztucznej inteligencji otworzyły nowe możliwości w zakresie wspierania i podtrzymywania pokoju i bezpieczeństwa – na przykład poprzez udoskonalenie technologii w dziedzinach takich jak wczesne ostrzeganie o konfliktach oraz weryfikacja kontroli zbrojeń i kontroli wywozu. Z drugiej strony, postępy te umożliwiły nowe sposoby tworzenia, lub zaostrzania, napięć, konfliktów i braku bezpieczeństwa pomiędzy państwami i wewnątrz nich. Zagrożenia, jakie stanowią niektóre sposoby stosowania sztucznej inteligencji – takie jak systemy śmiertelnej broni autonomicznej – okazały się poważną kwestią budzącą niepokój społeczności zajmującej się kontrolą zbrojeń. Jednym z aspektów tego zagrożenia – który zasługuje na większą uwagę, a który bieżące starania z zakresu kontroli zbrojeń i wysiłki dyplomatyczne mogą niedostatecznie uwzględniać – jest przekierowywanie i niewłaściwe stosowanie cywilnych badań i innowacji dotyczących sztucznej inteligencji przez nieodpowiedzialne podmioty, w tym podmioty niepaństwowe działające w złym zamiarze; obejmuje to niewłaściwe stosowanie generatywnych sieci przeciwstawnych do wytwarzania materiałów typu „deepfake” przeznaczonych do kampanii dezinformacyjnych.

Sztuczna inteligencja jest technologią prorozwojową o wielkim potencjale stosowania ogólnego. Badania i innowacje dotyczące sztucznej inteligencji opracowywane do zastosowań cywilnych mogą być (stosunkowo łatwo) pozyskiwane i wykorzystywane przez niektóre podmioty do szkodliwych lub wywrotowych celów, które mogą mieć skutki dla międzynarodowego pokoju i bezpieczeństwa. Przekierowywanie i niewłaściwe stosowanie technologii cywilnych nie są zjawiskiem nowym ani nie dotyczą wyłącznie sztucznej inteligencji. W pokrewnej dziedzinie robotyki zostało to niedawno uwidocznione przez uzbrojenie i zastosowanie rekreacyjnych dronów przez Daisz/ISIS w Syrii. Jednak w przypadku sztucznej inteligencji problem ten komplikuje wiele czynników: niematerialny i szybko zmieniający się charakter jej algorytmów i danych, co utrudnia kontrolę ich przenoszenia/rozprzestrzeniania; wiodąca rola sektora prywatnego w badaniach, rozwoju i ekosystemie innowacji, oraz wynikająca z tego potrzeba ochrony algorytmów własnościowych; oraz ogólnosiwiatowa dostępność wiedzy fachowej ludzi oraz zasobów materialnych zdolnych do zmiany celów technologii sztucznej inteligencji. Tymczasem osoby pracujące przy sztucznej inteligencji w sektorze cywilnym zbyt często pozostają nieświadome potencjalnych skutków, jakie dla międzynarodowego pokoju i bezpieczeństwa może mieć przekierowywanie i niewłaściwe stosowanie wyników ich pracy, względnie nie są one chętne do wzięcia udziału w bieżących dyskusjach na temat zagrożeń związanych ze sztuczną inteligencją, prowadzonych w środowiskach kontroli zbrojeń i nieprolifracji.

Istnieje potrzeba wspierania większego udziału cywilnej społeczności zajmującej się sztuczną inteligencją w pojmowaniu i łagodzeniu zagrożeń dla pokoju i bezpieczeństwa związanych z przekierowywaniem i niewłaściwym stosowaniem cywilnej technologii sztucznej inteligencji przez nieodpowiedzialne podmioty. W odniesieniu do Sztokholmskiego Międzynarodowego Instytutu Badań nad Pokojem (SIPRI) i Biura ONZ ds. Rozbrojenia (UNODA) może to zostać osiągnięte poprzez szersze propagowanie odpowiedzialnych innowacji w globalnej cywilnej społeczności zajmującej się sztuczną inteligencją. Poprzednie prace SIPRI i UNODA wykazały, że odpowiedzialne innowacje jako mechanizm samokontroli mogłyby zapewnić globalnej cywilnej społeczności zajmującej się sztuczną inteligencją praktyczne narzędzia i metody służące rozpoznaniu oraz pomocy w łagodzeniu zagrożeń, jakie przekierowywanie i niewłaściwe stosowanie cywilnych badań i innowacji dotyczących sztucznej inteligencji mogłoby stanowić dla pokoju i bezpieczeństwa, a także zapobieganiu tym zagrożeniom. W trakcie prac prowadzonych przez SIPRI i UNODA określono także metodologie i kilka bieżących ukierunkowanych na sferę cywilną inicjatyw dotyczących odpowiedzialnej sztucznej inteligencji, które można by wykorzystać do uwrażliwienia cywilnej społeczności zajmującej się sztuczną inteligencją na kwestie kontroli zbrojeń oraz nieprolifracji, a także na debaty ekspertów i stanowiska państw na temat odpowiedzialnego rozwijania, rozpowszechniania i stosowania sztucznej inteligencji, oraz doświadczenia, jakie można uzyskać z prac nad odpowiedzialnością w sektorze obrony¹. Co istotne, w trakcie tych poprzednich prac wyraźnie stwierdzono, że współdziałanie ze studentami nauk przyrodniczych, technologii, inżynierii i matematyki (STEM) – którzy jeszcze mają kontakt ze sztuczną inteligencją w środowisku edukacyjnym – jest bardzo istotne dla wszelkich skutecznych starań w zakresie odpowiedzialnych innowacji.

2. CELE

Projekty te mają na celu wspieranie większego udziału cywilnej społeczności zajmującej się sztuczną inteligencją w łagodzeniu zagrożeń, jakie przekierowywanie i niewłaściwe stosowanie cywilnych badań i innowacji dotyczących sztucznej inteligencji przez nieodpowiedzialne podmioty może stanowić dla międzynarodowego pokoju i bezpieczeństwa. Dokonuje się tego przez, po pierwsze, stworzenie lepszego pojmowania sposobu, w jaki rozwój i rozpowszechnianie badań i innowacji dotyczących sztucznej inteligencji może wpływać na zagrożenia wynikające z ich przekierowywania i niewłaściwego stosowania, oraz, po drugie, przez propagowanie odpowiedzialnych procesów, metod i narzędzi innowacji, co może pomóc w zapewnianiu pokojowego stosowania cywilnych innowacji i odpowiedzialnego rozpowszechniania wiedzy o sztucznej inteligencji. W tym celu w ramach tych projektów wspiera się działania z zakresu budowania zdolności, badań i zaangażowania, które i) zwiększą, za pośrednictwem procesów odpowiedzialnych innowacji, zdolności w ramach globalnej cywilnej społeczności zajmującej się sztuczną inteligencją do ujęcia i uwzględnienia zagrożeń dla pokoju i bezpieczeństwa, jakie stanowią przekierowywanie i niewłaściwe stosowanie cywilnej sztucznej inteligencji przez nieodpowiedzialne podmioty; oraz ii) wzmocnią powiązanie między staraniami na rzecz łagodzenia zagrożeń prowadzonymi w odniesieniu do odpowiedzialnej sztucznej inteligencji w sferze cywilnej a będącymi już w toku staraniami w ramach społeczności zajmującej się rozbrojeniem, kontrolą zbrojeń i nieprolifracją na szczeblu międzyrządowym. Co istotne, nie dążą one do stworzenia jakichkolwiek nowych standardów, zasad ani regulacji, ani do wkroczenia w dziedziny znajdujące się w kompetencji państw. Intencją tych projektów jest rozwijanie cywilnych starań w zakresie odpowiedzialnych innowacji, by objęły one zagrożenia dla pokoju i bezpieczeństwa wynikające z przekierowywania i niewłaściwego stosowania cywilnej sztucznej inteligencji przez nieodpowiedzialne podmioty, oraz zapewnianie informacji o istniejących staraniach międzyrządowych.

W celu skutecznego dotarcia do cywilnej społeczności zajmującej się sztuczną inteligencją i wywarcia wpływu na tę społeczność, w projektach stosuje się podejście trójtorowe, by:

¹ Metodyka obejmuje na przykład zalecane przez Instytut Inżynierów Elektryki i Elektroniki (IEEE) praktyki dotyczące oceny wpływu autonomicznych i inteligentnych systemów na ludzki dobrostan (norma IEEE 7010-220), czy opracowaną przez grupę ekspertów wysokiego szczebla ds. AI listę kontrolną oceny godnej zaufania sztucznej inteligencji. Inicjatywy obejmują: opracowaną przez IEEE globalną inicjatywę dotyczącą etyki dla autonomicznych i inteligentnych systemów; partnerstwo na rzecz sztucznej inteligencji; globalne partnerstwo na rzecz sztucznej inteligencji.

- (a) nawiązać kontakty z edukatorami – współpracować z wybranymi edukatorami i osobami tworzącymi akademickie programy nauczania nad opracowaniem i propagowaniem materiałów edukacyjnych, które mogą zostać wykorzystane do uwzględniania w programach kształcenia osób praktycznie zajmujących się sztuczną inteligencją kwestii dotyczących zagrożeń dla pokoju i bezpieczeństwa wynikających z przekierowywania i niewłaściwego stosowania badań i innowacji dotyczących cywilnej sztucznej inteligencji (np. w programach zajęć dotyczących etyki sztucznej inteligencji i odpowiedzialnych innowacji);
- (b) nawiązać kontakty ze studentami – zaznajomić wybranych studentów nauk przyrodniczych, technologii, inżynierii i matematyki z całego świata ze sposobami rozpoznawania zagrożeń dla pokoju i bezpieczeństwa wynikających z przekierowywania i niewłaściwego zastosowania cywilnej sztucznej inteligencji przez nieodpowiedzialne podmioty, i ze sposobami zapobiegania im lub ich złagodzenia w trakcie procesu badań i innowacji lub poprzez inne procesy zarządzania; oraz
- (c) nawiązać kontakty z branżą sztucznej inteligencji – współpracować ze stowarzyszeniami zawodowymi i organami ds. standardów takimi jak Instytut Inżynierów Elektryki i Elektroniki (IEEE) w celu i) rozpowszechniania wśród osób zawodowo zajmujących się techniką dostosowanych materiałów edukacyjnych oraz działań kontaktowych ; ii) wspierać pozytywne zastosowanie sztucznej inteligencji na rzecz pokoju i bezpieczeństwa; oraz iii) ułatwiać dialog i wymianę informacji między ekspertami ze środowisk akademickich, sektora prywatnego i sektora rządowego na temat sposobu załagodzenia zagrożenia wynikającego z przekierowywania i niewłaściwego stosowania cywilnych badań i innowacji dotyczących sztucznej inteligencji.

Takie podejście sprawia, że projekty dotrą na wszystkie poziomy społeczności zajmującej się sztuczną inteligencją, nie tylko do obecnych praktyków, lecz także do przyszłych pokoleń. Umożliwia ono także udział ze strony wszystkich sfer akademickich, branżowych i innych oraz wspiera trwałość przyszłych starań poprzez utworzenie sieci przekraczających granice tych sfer.

W projektach dąży się także do wykorzystywania możliwości i doświadczenia SIPRI i UNODA, by można było wywrzeć wpływ na społeczność zajmującą się sztuczną inteligencją na całym świecie, nie tylko na zainteresowane strony w UE. SIPRI i UNODA mają niepowtarzalne możliwości dotarcia do podmiotów zajmujących się sztuczną inteligencją w Afryce, w regionie Azji i Pacyfiku, w Europie oraz w Ameryce Północnej i Południowej oraz do umożliwiania kontaktów między tymi podmiotami. Obie instytucje mają także doświadczenie w pracy w innych dziedzinach nauki i technologii stojących w obliczu podobnych wyzwań z zakresu podwójnego zastosowania i rozprzestrzeniania, także w zakresie biotechnologii. W ramach projektów dąży się także do wykorzystania warunków występujących w Unii Europejskiej, takich jak a) istnienie zaawansowanych, obejmujących wiele podmiotów procesów dotyczących odpowiedzialnej sztucznej inteligencji; b) istniejący w UE wysoki poziom zaangażowania w kwestie rozbrojenia, kontroli zbrojeń oraz nieprolifracji, jak również wiedzy fachowej na te tematy; c) różnorodność powiązań, jakie organizacje akademickie, badawcze i sektora prywatnego w UE posiadają z innymi regionami, zwłaszcza na globalnym Południu, które także będzie stanowić istotny cel zaangażowania; oraz d) zróżnicowanie narodowościowe wśród studentów, edukatorów i inżynierów na uniwersytetach, w instytucjach badawczych i w sektorze prywatnym.

Włączenie wszystkich będzie stanowiło jeden z podstawowych aspektów przy prowadzeniu działań w ramach tych projektów. Aby skutecznie wspierać społeczność zajmującą się sztuczną inteligencją, w projektach dostrzega się fakt, że społeczność ta składa się z różnych podmiotów, a w szczególności że

- (a) aspekt płci jest wysoce istotnym czynnikiem. Z tego powodu aspekt płci będzie uwzględniany zgodnie z ogólnosystemowymi oenzetowskimi strategiami dotyczącymi uwzględniania aspektu płci i równości płci. Udział kobiet we wszystkich działaniach w ramach tego projektu będzie popierany i wymagany; oraz
- (b) w całości projektów będzie uwzględniane włączenie osób z niepełnosprawnościami oraz racjonalne zaspokajanie ich potrzeb. Będzie to obejmowało pokonywanie przeszkód stojących na drodze udziału

osób z niepełnosprawnościami, jak również zagwarantowanie, że zostaną podjęte kroki na rzecz współpracy z osobami z niepełnosprawnościami, aby ułatwić reprezentowanie merytorycznych poglądów i doświadczeń tych osób.

3. PROJEKTY

Trzy opisane poniżej projekty mają być komplementarne i wzajemnie się wspierać, a ich elementy będą realizowane przez 36 miesięcy.

Projekt 1 - Opracowanie materiałów edukacyjnych i budujących zdolności dla cywilnej społeczności zajmującej się sztuczną inteligencją

3.1.Cel projektu

Projekt 1 koncentruje się na dostarczeniu wiedzy i środków cywilnym podmiotom zajmującym się sztuczną inteligencją, aby ocenić i ograniczyć ryzyko, jakie dla międzynarodowego pokoju i bezpieczeństwa może stanowić przekierowywanie i niewłaściwe wykorzystywanie cywilnych badań i innowacji w zakresie sztucznej inteligencji przez nieodpowiedzialne podmioty. Jego celem jest opracowanie materiałów edukacyjnych i budujących zdolności które zapewnią praktykom w dziedzinie sztucznej inteligencji ze wszystkich regionów, poziomów i sektorów (w tym edukatorom skupionym na sztucznej inteligencji, twórcom programów nauczania, studentom STEM oraz inżynierom i badaczom sztucznej inteligencji w środowisku akademickim i sektorze prywatnym) informacje i narzędzia niezbędne do

- (a) zrozumienia, w jaki sposób cywilne badania i innowacje w zakresie sztucznej inteligencji mogą zostać przekierowane i niewłaściwie wykorzystane w sposób, który może stanowić ryzyko dla międzynarodowego pokoju i bezpieczeństwa, a także w jaki sposób decyzje dotyczące rozwoju i rozpowszechniania badań i innowacji mogą zwiększyć lub zmniejszyć ryzyko przekierowywania i niewłaściwego wykorzystywania;
- (b) zrozumienie wysiłków już podjętych przez społeczność zajmującą się rozbrojeniem, kontrolą zbrojeń i nieproliferacją w celu zmniejszenia ryzyka związanego z przekierowywaniem i niewłaściwym wykorzystywaniem cywilnych badań i innowacji; oraz
- (c) praktykowanie odpowiedzialnych innowacji w sposób ograniczający ryzyko przekierowywania i niewłaściwego wykorzystywania w rozwoju i rozpowszechnianiu badań i innowacji.

3.1.1. Opis projektu

W ramach tego projektu powstaną trzy oddzielne zestawy materiałów edukacyjnych i budujących zdolności.

- (a) Podręcznik (1) – W podręczniku zebrana zostanie podstawowa wiedza i środki dla podmiotów zajmujących się sztuczną inteligencją, które w procesie badań i innowacji pozwolą im ocenić i ograniczyć ryzyko przekierowywania i niewłaściwego wykorzystywania cywilnej technologii w zakresie sztucznej inteligencji przez nieodpowiedzialne podmioty. Omówione zostanie, dlaczego i w jaki sposób decyzje dotyczące rozwoju i rozpowszechniania badań i innowacji mogą wpływać na ryzyko przekierowywania i niewłaściwego wykorzystywania, a tym samym generować ryzyko lub szanse dla pokoju i bezpieczeństwa. Przedstawione zostaną również odpowiednie przepisy prawa międzynarodowego i zobowiązania w zakresie kontroli eksportu, a także względy bezpieczeństwa i ochrony, które są przedmiotem dyskusji w kręgach wojskowych, jak również w kręgach związanych z rozbrojeniem, kontrolą zbrojeń i nieproliferacją; oraz zaprezentowane przykładowe procesy i narzędzia do praktykowania odpowiedzialnych innowacji, takie jak metody oceny wpływu technologii i szablony oceny ryzyka.

- (b) Seria podcastów (~10) – Podcasty te będą stanowiły przystępne i zajmujące medium dla podmiotów zajmujących się sztuczną inteligencją, dzięki czemu uzyskają one wiedzę, dlaczego i w jaki sposób odpowiedzialne procesy innowacji w zakresie sztucznej inteligencji mogą wspierać międzynarodowy pokój i bezpieczeństwo poprzez ograniczanie ryzyka związanego z przekierowywaniem i niewłaściwym wykorzystywaniem przez nieodpowiedzialne podmioty. W ramach serii dokonany zostanie przegląd ważnych tematów (np. schemat przekierowywania i niewłaściwego wykorzystywania badań i innowacji podwójnego/ogólnego zastosowania; wyzwania humanitarne, strategiczne i polityczne związane z potencjalnym niewłaściwym wykorzystywaniem cywilnych innowacji w zakresie badań nad sztuczną inteligencją; wyzwania, przed którymi stoją środowiska związane z rozbrojeniem, kontrolą zbrojeń i nieprolifacją w ramach działań na rzecz ograniczania ryzyka; jak dokonywać odpowiedzialnych innowacji poprzez ocenę ryzyka; zgodność z przepisami dotyczącymi kontroli eksportu; zmniejszenie ryzyka na etapie projektowania; odpowiedzialne publikowanie; znajomość klientów; i doświadczenie w ćwiczeniach symulacyjnych), a osiłą poszczególnych podcastów będą wywiady, które zespół projektowy przeprowadzi z przedstawicielami odpowiednich społeczności.
- (c) Seria blogów (9-10) – Zespół opracuje serię wyselekcjonowanych postów na blogach, których celem będzie podniesienie rangi wysiłków zmierzających do przekroczenia granic między cywilnie ukierunkowaną "odpowiedzialną sztuczną inteligencją" a społecznościami zajmującymi się kontrolą zbrojeń i nieprolifacją. Seria blogów będzie stanowiła platformę do rozpowszechniania spostrzeżeń, pomysłów i rozwiązań dotyczących identyfikacji i przeciwdziałania ryzyku związanemu z przekierowywaniem i niewłaściwym wykorzystywaniem cywilnej sztucznej inteligencji w procesie badań i innowacji. Dołożone zostaną starania, by blog reprezentował różnorodność myśli i perspektyw obecnych w sektorze sztucznej inteligencji.

Materiały te będą rozpowszechniane publicznie poprzez strony internetowe podmiotów wdrażających, ich obecność w mediach społecznościowych oraz poprzez bezpośrednią komunikację z odpowiednimi jednostkami akademickimi, cywilnymi stowarzyszeniami zawodowymi w dziedzinie sztucznej inteligencji i innymi odpowiednimi grupami.

3.1.2. Oczekiwane rezultaty projektu

Oczekuje się, że projekt ten stworzy nowy zestaw materiałów, dzięki którym cywilni praktycy w dziedzinie sztucznej inteligencji będą mogli uwrażliwić się na to: a) w jaki sposób cywilne badania i innowacje w zakresie sztucznej inteligencji mogą zostać przekierowane i niewłaściwie wykorzystane w sposób, który może stanowić ryzyko dla międzynarodowego pokoju i bezpieczeństwa, b) w jaki sposób do takiego ryzyka podchodzi społeczność zajmująca się rozbrojeniem, kontrolą zbrojeń i nieprolifacją, oraz c) w jaki sposób praktycy w dziedzinie sztucznej inteligencji mogą dalej przyczyniać się do ograniczania takiego ryzyka poprzez odpowiedzialne procesy innowacyjne.

Oczekuje się, że przyczyni się to do zwiększenia zaangażowania cywilnego sektora sztucznej inteligencji w ograniczanie ryzyka, jakie dla międzynarodowego pokoju i bezpieczeństwa może stanowić przekierowywanie i niewłaściwe wykorzystywanie cywilnej sztucznej inteligencji; do poprawy zdolności osób zajmujących się zawodowo techniką do angażowania się w odpowiednie procesy w ramach społeczności zajmującej się rozbrojeniem, kontrolą zbrojeń i nieprolifacją, jak również do wspierania zaangażowania nowych odbiorców, którzy nie są tradycyjnie włączani w wysiłki edukacyjne w zakresie rozbrojenia i nieprolifacji.

Materiał ten ma również wspierać realizację pozostałych projektów i posłużyć jako podstawa do działań edukacyjnych i budujących zdolności w projekcie 2 oraz działań w zakresie dialogu i działań motywacyjnych w projekcie 3. Działania te mają z kolei wspomagać zwrotnie opracowywanie i udoskonalanie materiału. Oczekuje się, że takie iteratywne podejście pomoże rozwiązać potencjalne przeszkody w ich promowaniu, rozpowszechnianiu i stosowaniu w społeczności zajmującej się sztuczną inteligencją, w tym kwestie związane z językiem, treścią,

kontekstem i dostępnością, które mogłyby uniemożliwić ich oddziaływanie na poziomie globalnym, w szczególności w krajach globalnego Południa.

3.2. Projekt 2 – Działania w zakresie edukacji i budowania zdolności dla przyszłych praktyków w dziedzinie sztucznej inteligencji

3.2.1. Cel projektu

Celem projektu 2 jest wsparcie włączania problemu przekierowywania i niewłaściwego wykorzystywania cywilnych badań nad sztuczną inteligencją przez nieodpowiedzialne podmioty do edukacji przyszłych pokoleń praktyków w dziedzinie sztucznej inteligencji. Na dłuższą metę zagwarantuje to, że studenci nauk ścisłych, technologii, inżynierii i matematyki (STEM) kształtujący przyszłość sztucznej inteligencji będą świadomi negatywnego wpływu, jaki na międzynarodowy pokój i bezpieczeństwo mogą mieć przekierowywanie i niewłaściwe wykorzystywanie ich pracy przez nieodpowiedzialne podmioty, oraz że będą dysponować podstawowymi narzędziami niezbędnymi do zidentyfikowania i ograniczenia takiego ryzyka w procesie badań i innowacji.

W ramach tego projektu przeprowadzona zostanie seria warsztatów edukacyjnych i budujących zdolności z edukatorami i studentami we współpracy z wybranymi międzynarodowymi uczelniami i podmiotami branżowymi. Projekt ma zatem na celu opracowanie działań w zakresie budowania zdolności, które edukatorzy i twórcy akademickich programów nauczania mogliby wykorzystać do włączenia do szkolenia przyszłych praktyków w dziedzinie sztucznej inteligencji (np. kursy dotyczące etyki sztucznej inteligencji i odpowiedzialnych innowacji) oraz w zakresie rozważenia ryzyka związanego z przekierowywaniem i niewłaściwym wykorzystywaniem cywilnych badań nad sztuczną inteligencją i innowacji przez nieodpowiedzialne podmioty, a także powiązania ich z szerszym kontekstem pokoju i bezpieczeństwa. Poprzez te warsztaty w ramach projektu nastąpi próba utworzenia sieci zainteresowanych edukatorów, twórców programów nauczania i studentów, którzy wspieraliby rozpowszechnianie i promocję materiałów edukacyjnych projektu i działań związanych z budowaniem zdolności w społeczności edukacyjnej w zakresie sztucznej inteligencji i społeczności praktyków w dziedzinie sztucznej inteligencji. Ten element tworzenia sieci kontaktów ma na celu zapewnienie trwałego charakteru projektów po zakończeniu ich bezpośredniego okresu trwania oraz umożliwienie tworzenia silniejszych powiązań wspierających cywilne zaangażowanie techniczne na rzecz szerszych celów związanych z pokojem, bezpieczeństwem, rozbrojeniem i kontrolą zbrojeń.

3.2.2. Opis projektu

W ramach tego projektu zostanie przeprowadzona seria warsztatów edukacyjnych i budujących zdolności z edukatorami i studentami z wybranych uczelni z całego świata. Będą one stanowiły połączenie wykładów i interaktywnych działań, które zapewnią edukatorom i studentom możliwość zastanowienia się nad tym, w jaki sposób cywilne badania i innowacje w zakresie sztucznej inteligencji mogą zostać przekierowane i niewłaściwie wykorzystywane w sposób, który może stanowić ryzyko dla międzynarodowego pokoju i bezpieczeństwa, oraz w jaki sposób można zidentyfikować takie ryzyko, zapobiec mu lub ograniczyć je w procesie badań i innowacji lub poprzez inne procesy zarządcze. Działania te będą oparte na wcześniejszych pracach pilotażowych na mniejszą skalę przeprowadzonych przez UNODA, która wypróbowała metody współpracy ze studentami kierunków ścisłych i uwrażliwiania ich na znaczenie rozważenia szerszej pojmowanego oddziaływania ich pracy, jak również współpracy ze specjalistami spoza ich rodzimych dziedzin. Konkretnie byłyby to:

- (a) regionalne warsztaty budowania zdolności dla edukatorów i studentów (4) – w ich ramach będą prowadzone i promowane działania, które edukatorzy mogą wykorzystać do budowania zdolności studentów STEM w zakresie odpowiedzialnych innowacji w dziedzinie sztucznej inteligencji, ze szczególnym naciskiem na to, jak ocenić i ograniczyć ryzyko związane z przekierowywaniem i niewłaściwym wykorzystywaniem

cywilnej technologii w zakresie sztucznej inteligencji przez nieodpowiedzialne podmioty. Każdy warsztat będzie organizowany wspólnie z uniwersytetem z siedzibą w UE i prestiżowym uniwersytetem z innego regionu świata, dzięki czemu zawsze będzie łączył zróżnicowaną grupę uczestników z UE ze zróżnicowaną grupą uczestników spoza UE. Warsztaty obejmą następnie Amerykę Łacińską i Karaiby, Amerykę Północną, Afrykę oraz Azję i region Pacyfiku. Pozwoli to na udział studentów (na poziomie magisterskim i doktoranckim) z całego świata, w tym z globalnego Południa. Warsztaty byłyby prowadzone głównie w języku angielskim, ale tam, gdzie to możliwe, uczestnicy mieliby możliwość zaangażowania się w działania oparte na alternatywnych grupach językowych; oraz

- (b) międzynarodowe warsztaty na temat zrównoważonego budowania zdolności (1) – w ich ramach zostaną wykorzystane wnioski wyciągnięte z warsztatów regionalnych, zostanie ułatwiona wymiana informacji i doświadczeń pomiędzy edukatorami i wybranymi studentami z uczelni zaangażowanych w projekt. Podczas warsztatów omówione zostaną sposoby dopracowania działań i narzędzi wypracowanych w trakcie realizacji projektu oraz upowszechnienia ich poza grupą uczestniczących uczelni. Omówione zostaną również sposoby wspierania zaangażowania studentów – po wejściu przez nich na rynek pracy – na rzecz odpowiedzialnej sztucznej inteligencji, która przeciwdziała ryzyku dla międzynarodowego pokoju i bezpieczeństwa wynikającym z przekierowywania i niewłaściwego wykorzystywania.

Sieci i obecność SIPRI i UNODA w Afryce, regionie Azji i Pacyfiku, Europie, Ameryce Północnej i Południowej zostaną wykorzystane do ułatwienia i wspierania niektórych aspektów działań, w zależności od potrzeb.

3.2.3. Oczekiwane rezultaty projektu

Oczekuje się, że projekt stworzy modele działań w zakresie budowania zdolności i zaangażowania, które edukatorzy i twórcy programów akademickich mogliby powielać, aby uwrażliwić przyszłych praktyków w dziedzinie sztucznej inteligencji na problemy związane z przekierowywaniem i niewłaściwym wykorzystywaniem cywilnej sztucznej inteligencji przez nieodpowiedzialne podmioty oraz na to, jak mogą pomóc w łagodzeniu tych problemów poprzez odpowiedzialne procesy innowacyjne. Oczekuje się, że po zakończeniu działań projektowych uczestnicy (edukatorzy, ale także studenci STEM) będą potrafili wykorzystywać i promować narzędzia, metody i koncepcje odpowiedzialnych innowacji w celu identyfikacji i ograniczania ryzyka przekierowywania i niewłaściwego wykorzystywania w rozwoju i upowszechnianiu cywilnych badań i innowacji w zakresie sztucznej inteligencji.

Oczekuje się również, że rezultatem działań w ramach projektu będzie stworzenie sieci edukatorów, twórców programów nauczania i studentów, którzy nie tylko będą promować działania projektowe w ramach społeczności zajmujących się edukacją w zakresie sztucznej inteligencji i środowisk zawodowych (np. podczas konferencji IEEE Computational Intelligence Society), ale również będą w stanie wnieść swój techniczny wkład w międzynarodowe procesy zarządcze kierowane przez państwo (np. proces Konwencji o Pewnych Broniach Konwencjonalnych dotyczący powstających technologii w obszarze systemów śmiertelnej broni autonomicznej).

Krótko- i długoterminowa wartość tych działań zostanie wykazana przez ankiety przed i po zakończeniu działań.

3.3. Projekt 3 – Ułatwianie zrównoważonego rozwoju na dłuższą metę, upowszechniania i wpływu odpowiedzialnych innowacji w zakresie sztucznej inteligencji na pokój i bezpieczeństwo

3.3.1. Cel projektu

Celem projektu 3 jest ułatwienie zrównoważonego rozwoju na dłuższą metę, upowszechniania i wpływu odpowiedzialnych innowacji w zakresie sztucznej inteligencji jako środka ograniczania ryzyka, jakie dla pokoju i bezpieczeństwa może stanowić przekierowywanie i niewłaściwe wykorzystywanie cywilnych badań i innowacji w zakresie sztucznej inteligencji. Osiągnięcie tego celu ma nastąpić poprzez organizowanie okrągłych stołów z przedstawicielami branży sztucznej inteligencji, wielostronne dialogi, stworzenie publicznego sprawozdania oraz ukierunkowane działania popularyzatorskie. Projekt ma na celu zagwarantowanie, że wyniki tych prac, w szczególności w zakresie edukacji, budowania potencjału i działań na rzecz zaangażowania, dotrą do całej społeczności zajmującej się sztuczną inteligencją i wpłyną na nią na wszystkich poziomach (od studentów do inżynierów i innych specjalistów w zakresie sztucznej inteligencji) i ponad granicami geograficznymi, sektorowymi i innymi. Aby zwiększyć możliwość wywarcia szerokiego i głębokiego wpływu, konieczna jest współpraca z profesjonalnymi organizacjami działającymi w tej przestrzeni, takimi jak IEEE, oraz prowadzenie wielowymiarowych działań na rzecz zaangażowania w sferach akademickich, branżowych i innych. Takie działania dadzą zainteresowanym przedstawicielom różnych społeczności zajmujących się sztuczną inteligencją możliwość wzięcia odpowiedzialności za problem i przedstawienia własnych poglądów na temat tego, w jaki sposób wysiłki zmierzające do ograniczenia ryzyka mogą być prowadzone i promowane w sposób zrównoważony w obrębie globalnej społeczności zajmującej się sztuczną inteligencją. Aby projekt okazał się wartościowy w dłuższej perspektywie dla państw, organizacji międzyrządowych i innych, ważne jest również to, aby praktycy w dziedzinie sztucznej inteligencji mogli uczyć się od ekspertów rządowych zaangażowanych w ograniczanie ryzyka w kontekście rozbrojenia, kontroli zbrojeń i nieprolifracji, a także współpracować z nimi. Kluczowe znaczenie dla trwałości projektu ma również zagwarantowanie, by spostrzeżenia zebrane w wyniku działań angażujących były przeanalizowane, skonsolidowane i odpowiednio upowszechnione.

3.3.2. Opis projektu

Na projekt składają się następujące kluczowe wątki:

- (a) wielostronne dialogi na temat „odpowiedzialnych innowacji w zakresie sztucznej inteligencji na rzecz pokoju i bezpieczeństwa” (do 9) – ta seria wirtualnych spotkań dialogowych zgromadziłaby ekspertów ze środowisk akademickich, badawczych, sektora prywatnego i tradycyjnej kontroli zbrojeń z UE i spoza niej w celu omówienia
 - i. tendencji technologicznych, które mogą powodować ryzyko rozprzestrzeniania, przekierowywania i niewłaściwego wykorzystywania mające wpływ na pokój i bezpieczeństwo międzynarodowe;
 - ii. sposobów angażowania się w ograniczanie ryzyka poprzez odpowiedzialne procesy, metody i środki innowacyjne oraz możliwości i wyzwania związane z dialogiem i dzieleniem się wiedzą w gronie społeczności zainteresowanych stron, z uwzględnieniem działających w innych sektorach, takich jak przemysł biologiczny i chemiczny; oraz
 - iii. potencjalnej wartości, celu i formatu samowystarczальной sieci ekspertów i działań w zakresie dialogu. Grupa ekspertów będzie spotykać się kilka razy w roku i pracować nad zorganizowaniem dwóch publicznych wydarzeń dla szerszej społeczności.

Z dziewięciu wirtualnych spotkań dwa mają być otwarte dla publiczności, aby ułatwić szersze konsultacje.

- (b) okrągłe stoły dla sektora prywatnego (do 6) – ta seria wirtualnych okrągłych stołów zainicjuje dialog z podmiotami zajmującymi się odpowiedzialnymi procesami innowacyjnymi w zakresie sztucznej inteligencji w sektorze prywatnym (np. partnerstwa w zakresie sztucznej inteligencji) na temat tego, w jaki sposób mogą one przyczynić się do zminimalizowania ryzyka związanego z przekierowywaniem i niewłaściwym wykorzystywaniem technologii cywilnej sztucznej inteligencji przez nieodpowiedzialne podmioty, a także zbadać ewentualne zachęty w ramach rozwoju sektora prywatnego do takiego działania. Wśród tematów znajdują się
- znaczenie kontekstu bezpieczeństwa międzynarodowego i rozbrojenia dla sektora prywatnego;
 - środowisko(-a) prawne, w którym(-ych) sztuczna inteligencja jest rozwijana, wdrażana i eksploatowana na całym świecie;
 - kwestia, w jaki sposób wykorzystać mechanizmy oceny ryzyka i inne środki, które są częścią procesów odpowiedzialnej innowacji i programów zapewniania zgodności przepisów dotyczących działalności przedsiębiorstw lub mogą być do tych procesów i programów włączone; oraz
 - wnioski, jakie można wyciągnąć z innych gałęzi przemysłu, procesów i ram odnoszących się do kontroli zbrojeń (np. przemysł biologiczny i chemiczny).
- (c) sprawozdanie na temat perspektyw społeczności zajmującej się sztuczną inteligencją w zakresie kontroli zbrojeń i ograniczania ryzyka w sztucznej inteligencji, skierowane do środowisk zajmujących się sztuczną inteligencją i kontrolą zbrojeń (I) – opracowanie tego sprawozdania pozwoli uchwycić i skonsolidować ustalenia i zalecenia projektu w jednym dokumencie referencyjnym skierowanym zarówno do odpowiedzialnych społeczności zajmujących się cywilną sztuczną inteligencją, jak i do społeczności zajmujących się kontrolą zbrojeń. W sprawozdaniu zostanie omówione, w jaki sposób można zidentyfikować, ocenić i rozwiązać międzynarodowe ryzyko dla pokoju i bezpieczeństwa związane z przekierowywaniem i niewłaściwym wykorzystywaniem cywilnych badań i innowacji w zakresie sztucznej inteligencji.
- (d) wydarzenia związane z upowszechnianiem, ukierunkowane na konsultacje i zaangażowanie społeczności zajmujących się sztuczną inteligencją i kontrolą zbrojeń (do uzgodnienia) – zespół będzie szukał możliwości przekazania informacji o pracach, ich wynikach i zaleceniach w trakcie trwania projektu. Format wydarzeń i treść prezentacji będą dostosowane do potrzeb grup docelowych. Mogą to być spotkania CONOP, Europejskiego Zgromadzenia Sojuszu AI; grupy ekspertów rządowych ds. nowych technologii w dziedzinie śmiertelnych autonomicznych systemów broni; Międzyagencyjnej Grupy Roboczej ds. Sztucznej Inteligencji (IAWG-AI); inicjatywy Międzynarodowego Związku Telekomunikacyjnego *AI for Good*; Dorocznego Dialogu Innowacyjnego UNIDIR; oraz Instytutu Inżynierów Elektryki i Elektroniki. Zespół będzie również dążył do nawiązania dwustronnej współpracy z odpowiednimi zainteresowanymi stronami z rządu, środowisk akademickich i sektora prywatnego.

3.3.3. Spodziewane wyniki projektu

Oczekuje się, że projekt ten położy podwaliny pod zrównoważony rozwój, upowszechnianie i wpływ odpowiedzialnych innowacji w procesach związanych ze sztuczną inteligencją dotyczących ryzyka związanego z rozprzestrzenianiem, przekierowywaniem i niewłaściwym wykorzystywaniem oraz skutków tego ryzyka dla pokoju i bezpieczeństwa wykraczających poza bezpośredni okres obowiązywania decyzji Rady.

Oczekuje się, że wielostronny dialog zapewni model wymiany informacji i współpracy w zakresie ograniczania ryzyka nie tylko w ramach globalnej społeczności zajmującej się sztuczną inteligencją, ale także pomiędzy odpowiedzialną społecznością zajmującą się cywilną

sztuczną inteligencją a społecznościami zajmującymi się rozbrojeniem, kontrolą zbrojeń i nieproliferacją. Taki model mógłby zostać wykorzystany do zaznajomienia decydentów z kluczowymi osiągnięciami technologicznymi i naukowymi istotnymi dla odpowiedzialnych innowacji w zakresie sztucznej inteligencji, a także do zapoznania odbiorców technicznych ze środowiskiem, w które obecnie angażują się decydenci. Oczekuje się, że projekt ułatwi trwałe relacje i zaangażowanie pomiędzy zainteresowanymi podmiotami w obrębie tych różnych społeczności i pomiędzy nimi. Oczekuje się, że takie heterogeniczne efekty sieciowe umożliwią większy rozwój i powszechną promocję odpowiedzialnych innowacji w dziedzinie sztucznej inteligencji na rzecz pokoju i bezpieczeństwa poza ramami czasowymi projektu.

Oczekuje się, że dialog sektora prywatnego umożliwi większe i głębsze zaangażowanie sektora prywatnego zajmującego się sztuczną inteligencją w identyfikację, zapobieganie i ograniczanie ryzyka dla pokoju i bezpieczeństwa wynikającego z przekierowywania i niewłaściwego wykorzystywania cywilnych badań i innowacji w zakresie sztucznej inteligencji. Oczekuje się, że projekt zapewni kluczowym uczestnikom procesów w sektorze prywatnym większe zrozumienie i poczucie odpowiedzialności za problemy, które stara się rozwiązać. Ponadto ma on na celu ułatwienie (szerszego) przyjęcia i wdrożenia procesów, metod i środków odpowiedzialnych innowacji w istniejących mechanizmach i procedurach zarządzania ryzykiem korporacyjnym.

Oczekuje się, że wielostronny dialog i okrągły stół sektora prywatnego przyniosą również spostrzeżenia dotyczące szeregu istotnych kwestii, w tym a) sposobu dalszego dopracowania i zastosowania metod i środków odpowiedzialnych innowacji w celu identyfikacji, zapobiegania i ograniczania ryzyka wynikającego z przekierowywania i niewłaściwego wykorzystywania cywilnych badań i innowacji w zakresie sztucznej inteligencji; b) sposobu pozytywnego wykorzystywania badań i innowacji w zakresie sztucznej inteligencji, aby wspierać cele związane z pokojem i bezpieczeństwem (np. zastosowania we wczesnym ostrzeganiu przed konfliktami i w pomocy humanitarnej); oraz c) sposobu ułatwienia szerszego dialogu i wymiany informacji między działaniami na rzecz ograniczania ryzyka podejmowanymi przez odpowiedzialną społeczność zajmującą się cywilną sztuczną inteligencją (np. różne inicjatywy prowadzone przez IEEE) a wysiłkami podejmowanymi już przez społeczność zajmującą się rozbrojeniem, kontrolą zbrojeń i nieproliferacją na szczeblu międzyrządowym.

W ramach sprawozdania i działań w zakresie upowszechniania nastąpi analiza, konsolidacja i upowszechnienie spostrzeżeń uzyskanych w ramach projektów 1, 2 i 3, co będzie skutkowało wsparciem promowania ustaleń i zaleceń wynikających z działań projektowych w obrębie globalnej społeczności zajmującej się sztuczną inteligencją, jak również w ramach społeczności zaangażowanej w politykę w tym zakresie. Oczekuje się również, że pomogą one w zapewnieniu trwałości wpływu po upływie okresu realizacji projektów.

4. CZAS TRWANIA

Całkowity szacowany czas realizacji projektów wynosi 36 miesięcy.