



Conselho da  
União Europeia

Bruxelas, 26 de outubro de 2022  
(OR. en)

14095/22  
ADD 1

LIMITE

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

## PROPOSTA

---

de:	Alto representante da União para os Negócios Estrangeiros e a Política de Segurança, com a assinatura de Stefano SANNINO, secretário-geral
data de receção:	26 de outubro de 2022
para:	Thérèse Blanchet, secretária-geral do Conselho da União Europeia
Assunto:	Proposta de decisão do Conselho de apoio à realização do projeto "Promover a inovação responsável no domínio da inteligência artificial para a paz e a segurança", apresentada ao Conselho pelo alto representante da União para os Negócios Estrangeiros e a Política de Segurança

---

Envia-se em anexo, à atenção das delegações, o documento HR(2022) 238.

---

Anexo: HR(2022) 238

# SERVIÇO EUROPEU PARA A AÇÃO EXTERNA



## DOCUMENTO DO PROJETO

**Promover a inovação responsável no domínio da inteligência artificial para a paz e a segurança**

**HR(2022)238**

### 1. CONTEXTO

Os recentes progressos no domínio da inteligência artificial (IA) desencadearam novas possibilidades para apoiar e incentivar a paz e a segurança, por exemplo, através de melhorias tecnológicas em domínios como o alerta precoce de conflitos e a verificação do controlo de armas e das exportações. Por outro lado, estes progressos permitiram novos meios para gerar ou agravar tensões, conflitos e a insegurança entre os Estados e no interior destes. Os riscos apresentados por determinadas aplicações da IA, como os sistemas de armas letais autónomos, surgiram como uma das principais preocupações da comunidade do controlo de armas. Uma via de risco que merece maior atenção e para a qual os atuais esforços diplomáticos e de controlo de armas podem ser insuficientes para lhe dar resposta é o desvio e a utilização indevida da investigação e inovação civis no domínio da IA por parte de intervenientes irresponsáveis, incluindo intervenientes não estatais mal-intencionados, como é o caso da utilização abusiva de redes generativas antagónicas para produzir falsificações profundas para campanhas de desinformação.

A IA é uma tecnologia facilitadora com grande potencial de utilização geral. A investigação e a inovação no domínio da IA desenvolvidas para aplicações civis podem ser acedidas (de forma relativamente fácil) e reorientadas por determinados intervenientes para utilizações nocivas ou disruptivas que podem ter implicações para a paz e a segurança internacionais. O desvio e a utilização indevida de tecnologias civis não são fenómenos novos nem são exclusivos da IA. No domínio conexo da robótica, esta situação foi recentemente ilustrada pela utilização de drones recreativos como armas pelo Daexe/EIIL na Síria. No entanto, no caso da IA, o problema é agravado por múltiplos fatores: a natureza intangível e em rápida evolução dos algoritmos e dados da IA, o que dificulta o controlo da transferência/proliferação desses algoritmos e dados; o papel de liderança do setor privado no ecossistema da investigação, desenvolvimento e inovação e a conseqüente necessidade de proteger a propriedade intelectual desses algoritmos; e a disponibilidade a nível mundial das competências humanas e dos recursos materiais capazes de reorientar as tecnologias da IA. Entretanto, as pessoas que trabalham na IA no setor civil continuam a desconhecer, com demasiada frequência, as potenciais implicações que o desvio e a utilização indevida do seu trabalho podem ter para a paz e a segurança internacionais ou hesitam em participar nos debates existentes sobre os riscos da IA nos círculos do controlo de armas e da não proliferação.

É necessário apoiar uma maior participação da comunidade civil da IA na compreensão e atenuação dos riscos para a paz e a segurança associados ao desvio e à utilização indevida da tecnologia civil da IA por intervenientes irresponsáveis. Para o Instituto Internacional de Estocolmo para a Investigação sobre a Paz (SIPRI) e o Gabinete das Nações Unidas para os Assuntos de Desarmamento (GNUAD), tal poderia ser alcançado através de uma maior promoção da inovação responsável na comunidade civil mundial da IA. O trabalho realizado pelo SIPRI e pelo GNUAD demonstrou que a inovação responsável, enquanto mecanismo de autogovernança, pode fornecer à comunidade civil mundial da IA instrumentos e métodos práticos para identificar e ajudar a prevenir e atenuar os riscos que o desvio e a utilização indevida da investigação e inovação civil no domínio da IA podem representar para a paz e a segurança. O trabalho do SIPRI e do GNUAD também identificou metodologias e várias iniciativas em curso centradas em civis sobre IA responsável que poderiam ser desenvolvidas para sensibilizar a comunidade civil da IA para as questões do controlo de armas e da não proliferação, realizar debates de peritos e definir as posições dos Estados sobre o desenvolvimento responsável, a difusão e a utilização da IA, bem como para retirar ensinamentos do trabalho de responsabilidade do setor da defesa<sup>1</sup>. Fundamentalmente, este trabalho identificou claramente a colaboração com os estudantes da ciência, tecnologia, engenharia e matemática (CTEM), que ainda estão a trabalhar com a IA num formato educativo, como um elemento central de qualquer esforço de inovação eficaz e responsável.

## 2. OBJETIVOS

Estes projetos visam apoiar uma maior participação da comunidade civil da IA na atenuação dos riscos que o desvio e a utilização indevida da investigação e inovação civis no domínio da IA por intervenientes irresponsáveis podem representar para a paz e a segurança internacionais. Visam fazê-lo, em primeiro lugar, ao gerar uma maior compreensão da forma como as decisões no desenvolvimento e difusão da investigação e inovação no domínio da IA podem afetar os riscos de desvio e de utilização indevida e, por sua vez, gerar riscos ou oportunidades para a paz e a segurança e, em segundo lugar, ao promover processos, métodos e instrumentos de inovação responsáveis que possam ajudar a assegurar a aplicação pacífica das inovações civis e a difusão responsável dos conhecimentos em matéria de IA. Para o efeito, apoiam o reforço das capacidades, da investigação e das atividades de participação que *i*) reforçarão a capacidade da comunidade civil mundial da IA para incluir e enfrentar os riscos para a paz e a segurança decorrentes do desvio e da utilização indevida da IA civil por intervenientes irresponsáveis através de processos de inovação responsáveis; e *ii*) consolidarão a ligação entre os esforços de atenuação dos riscos em matéria de IA responsável na esfera civil e os esforços já em curso na comunidade do desarmamento, do controlo de armas e da não proliferação a nível intergovernamental. Fundamentalmente, não tencionam estabelecer novas normas, princípios ou regulamentação, nem intervir em domínios da competência dos Estados. Em vez disso, tencionam desenvolver esforços civis de inovação responsáveis para incluir os riscos para a paz e a segurança decorrentes do desvio e da utilização indevida da IA civil por intervenientes irresponsáveis e proporcionar educação sobre os esforços intergovernamentais pertinentes em curso.

---

<sup>1</sup> As metodologias incluem, por exemplo, as práticas recomendadas pelo Instituto de Engenharia Eletrotécnica e Eletrónica (IEEE) para avaliar o impacto dos sistemas autónomos e inteligentes no bem-estar humano (IEEE Std 7010-2020) ou a lista de avaliação para uma inteligência artificial de confiança (ALTAI) do Grupo de Peritos de Alto Nível em Inteligência Artificial. As iniciativas incluem: a Iniciativa Global IEEE para Considerações Éticas em Inteligência Artificial e Sistemas Autónomos; a parceria para a inteligência artificial; a Parceria Global para a Inteligência Artificial.

Para alcançar e influenciar eficazmente a comunidade civil da IA, os projetos adotam uma abordagem em três vertentes, procurando

- (a) colaborar com os educadores – trabalhar com educadores e responsáveis pela elaboração dos programas de estudos académicos selecionados no desenvolvimento e na promoção de materiais educativos que possam ser utilizados para integrar a consideração dos riscos para a paz e a segurança decorrentes do desvio e da utilização indevida da investigação e inovação civil no domínio da IA por intervenientes irresponsáveis na formação de futuros profissionais da IA (por exemplo, em cursos sobre ética no domínio da IA e inovação responsável);
- (b) dialogar com os estudantes – demonstrar a estudantes selecionados nos domínios da ciência, tecnologia, engenharia e matemática (CTEM) de todo o mundo a forma como os riscos para a paz e a segurança decorrentes do desvio e da utilização indevida do desenvolvimento civil da IA por intervenientes irresponsáveis podem ser identificados, evitados ou atenuados no processo de investigação e inovação ou através de outros processos de governação; e
- (c) colaborar com a indústria da IA – trabalhar com associações profissionais e organismos de normalização, como o Instituto de Engenharia Eletrotécnica e Eletrónica (IEEE), para i) divulgar materiais educativos adaptados e atividades de participação junto dos profissionais técnicos; ii) apoiar as utilizações positivas da IA para a paz e a segurança; e iii) facilitar o diálogo e a partilha de informações entre peritos do meio académico, do setor privado e do governo sobre a forma de atenuar o risco de desvio e utilização indevida da investigação e inovação civil no domínio da IA por intervenientes irresponsáveis.

Esta abordagem permite que os projetos cheguem à comunidade da IA a todos os níveis, incluindo não só os atuais profissionais, mas também as gerações futuras. Permite também a participação de todo o mundo académico, da indústria e de outros grupos, e apoia a sustentabilidade dos esforços futuros através da criação de redes que atravessem essas fronteiras.

Os projetos procuram também utilizar o poder agregador e a experiência do SIPRI e do GNUAD para influenciar a comunidade da IA a nível mundial e não apenas as partes interessadas da UE. O SIPRI e o GNUAD encontram-se numa posição privilegiada para alcançar e facilitar o diálogo entre os intervenientes no domínio da IA de África, Ásia-Pacífico, Europa e América do Norte e do Sul. Ambas as entidades têm também experiência noutros domínios científicos e tecnológicos que enfrentam desafios semelhantes em matéria de dupla utilização e proliferação, incluindo a biotecnologia. Os projetos procuram igualmente tirar partido das condições existentes na União Europeia, tais como a) a existência de processos avançados e multilaterais em matéria de IA responsável; b) o elevado nível de envolvimento e de conhecimentos especializados em questões de desarmamento, controlo de armas e não proliferação na UE; c) a diversidade de ligações que as organizações académicas, de investigação e do setor privado da UE têm com outras regiões, nomeadamente no hemisfério sul, o que constituirá também um importante objetivo de participação; e d) a diversidade das nacionalidades dos estudantes, educadores e engenheiros das universidades, dos institutos de investigação e do setor privado.

A inclusão será uma consideração fundamental para a realização das atividades dos projetos. A fim de apoiar eficazmente a comunidade da IA, os projetos reconhecem que a comunidade da IA é constituída por um leque diversificado de intervenientes e, em particular, que

- (a) a perspectiva de género é um fator altamente relevante. Por este motivo, as questões de género serão integradas em consonância com as estratégias de integração da perspectiva de género e de paridade a nível do sistema das Nações Unidas. A participação das mulheres em todas as atividades do projeto será incentivada e exigida; e
- (b) a inclusão das pessoas com deficiência e as adaptações razoáveis a necessidades serão realizadas ao longo de todo o processo. Tal incluirá a eliminação dos obstáculos à participação das pessoas com deficiência, bem como a garantia de que são tomadas medidas para envolver e facilitar a representação dos pontos de vista e experiências significativas das pessoas com deficiência relativamente a questões de lhes interessam.

### **3. PROJETOS**

Os três projetos a seguir descritos destinam-se a ser complementares e a apoiar-se mutuamente, com elementos a decorrer ao longo dos 36 meses.

#### **Projeto 1 – Produção de material educativo e de reforço de capacidades para a comunidade civil da IA**

##### **3.1. Objetivo do projeto**

O projeto 1 centra-se na disponibilização de conhecimentos e meios para que os intervenientes civis no domínio da IA avaliem e atenuem os riscos que o desvio e a utilização indevida da investigação e inovação civis no domínio da IA por intervenientes irresponsáveis podem representar para a paz e a segurança internacionais. Visa produzir material educativo e de reforço das capacidades que forneça aos profissionais da IA de todas as regiões, níveis e setores (incluindo educadores centrados na IA, responsáveis pela elaboração dos programas de estudos, estudantes CTEM e engenheiros e investigadores da IA do meio académico e do setor privado) as informações e os instrumentos necessários para

- (a) compreender de que modo a investigação e a inovação civis no domínio da IA podem ser desviadas e indevidamente utilizadas de formas que possam representar riscos para a paz e a segurança internacionais e de que modo as decisões em matéria de desenvolvimento e difusão da investigação e inovação podem aumentar ou diminuir o risco de desvio e utilização indevida;
- (b) compreender os esforços já empreendidos pela comunidade do desarmamento, do controlo de armas e da não proliferação para atenuar os riscos de desvio e utilização indevida da investigação e inovação civis; e
- (c) praticar uma inovação responsável de forma a atenuar o risco de desvio e utilização indevida no desenvolvimento e difusão da investigação e inovação.

##### **3.1.1. Descrição do projeto**

Este projeto produzirá três conjuntos independentes de materiais de educação e desenvolvimento de capacidades.

- (a) *Manual (1)* – O manual compilará conhecimentos básicos e meios para os intervenientes no domínio da IA avaliarem e atenuarem, no processo de investigação e inovação, os riscos de desvio e utilização indevida da tecnologia da IA civil por intervenientes irresponsáveis. Debaterá por que motivo e de que forma as decisões sobre o desenvolvimento e a difusão da investigação e da inovação podem afetar os riscos de desvio e de utilização indevida e, por sua vez, gerar riscos ou oportunidades para a paz e a segurança. O manual incluirá igualmente as obrigações pertinentes do direito internacional e de controlo das exportações, para além das considerações de segurança e proteção em debate nos círculos militares, do desarmamento, do controlo de armas e da não proliferação; e apresentará exemplos de processos e ferramentas para a prática de inovação responsável, tais como metodologias de avaliação do impacto tecnológico e modelos de avaliação dos riscos.
- (b) *Série de podcasts (~10)* – Estes podcasts funcionarão como um meio acessível e atraente para que os intervenientes na IA saibam por que motivo e de que forma os processos de inovação no domínio da IA responsável podem apoiar a paz e a segurança internacionais através da atenuação dos riscos apresentados pelo desvio e pela utilização indevida de intervenientes irresponsáveis. A série de *podcasts* analisará temas importantes (por exemplo, o padrão de desvio e de utilização indevida da investigação e inovação de utilização dupla/geral; os desafios humanitários, estratégicos e políticos associados à potencial utilização indevida da investigação e inovação civil no domínio da IA; os desafios que os círculos do desarmamento, do controlo de armas e da não proliferação enfrentam nos esforços de atenuação dos riscos; como investigar de forma responsável, através da avaliação dos riscos; a conformidade com o controlo das exportações; a redução dos riscos desde a conceção; a publicação responsável; conhecer os seus clientes; e experiência de exercícios teóricos) e será estruturada em torno de entrevistas que a equipa do projeto realizará a representantes das comunidades pertinentes.
- (c) *Série de blogues (9-10)* – A equipa desenvolverá uma série de publicações selecionadas de blogues, destinada a aumentar a visibilidade dos esforços que tentam atravessar os limites entre a "IA responsável" centrada em civis e as comunidades do controlo de armas e da não proliferação. A série de blogues proporcionará uma plataforma para divulgar perspetivas, ideias e soluções sobre a identificação e a abordagem dos riscos associados ao desvio e à utilização indevida da IA civil no processo de investigação e inovação. O blogue procurará representar a diversidade de pensamentos e perspetivas presentes no setor da IA.

Estes materiais serão divulgados publicamente através dos sítios Web dos intervenientes responsáveis pela execução, da sua presença nas redes sociais e da comunicação direta com as entidades académicas e as associações profissionais civis pertinentes no domínio da IA, bem como com outros grupos adequados.

### **3.1.2. Resultados esperados do projeto**

Espera-se que este projeto estabeleça um novo conjunto de materiais através dos quais os profissionais civis da IA possam ser sensibilizados para a) a forma como a investigação e a inovação civis no domínio da IA podem ser desviadas e indevidamente utilizadas de formas que possam representar riscos para a paz e a segurança internacionais, b) como esses riscos estão a ser abordados pela comunidade do desarmamento, do controlo de armas e da não proliferação, e c) como os profissionais da IA podem contribuir ainda mais para a atenuação desses riscos através de processos de inovação responsáveis.

Espera-se que isto impulse a participação do setor civil da IA na atenuação dos riscos que o desvio e a utilização indevida da IA podem representar para a paz e a segurança internacionais; melhore a capacidade dos profissionais técnicos de participarem em processos pertinentes na comunidade do desarmamento, do controlo de armas e da não proliferação, bem como apoie a participação de novos públicos tradicionalmente não incluídos nos esforços de desarmamento e de educação para a não proliferação.

O material deverá também apoiar a execução dos outros projetos e servirá de base para as atividades educativas e de reforço das capacidades no âmbito do projeto 2, bem como para as atividades de diálogo e participação no projeto 3. Estas atividades deverão, por sua vez, contribuir para a produção e o aperfeiçoamento do material. Espera-se que essa abordagem iterativa ajude a eliminar os potenciais obstáculos à sua promoção, difusão e utilização na comunidade da IA, incluindo questões relacionadas com a língua, os conteúdos, o contexto e a disponibilidade, que possam impedir o seu impacto a nível mundial, em especial no hemisfério sul.

### **3.2. Projeto 2 – Atividades de educação e reforço de capacidades para futuros profissionais da IA**

#### **3.2.1. Objetivo do projeto**

O objetivo do projeto 2 é apoiar a integração do problema do desvio e da utilização indevida da investigação civil no domínio da IA por intervenientes irresponsáveis na educação das futuras gerações de profissionais da IA. A longo prazo, tal assegurará que os estudantes em ciência, tecnologia, engenharia e matemática (CTEM) que moldam o futuro da IA estão conscientes dos impactos negativos que o desvio e a utilização indevida do seu trabalho por intervenientes irresponsáveis podem ter na paz e segurança internacionais e que disporão dos instrumentos básicos necessários para identificar e atenuar esse risco no processo de investigação e inovação.

Este projeto organizará uma série de seminários educativos e de reforço das capacidades com educadores e estudantes, em colaboração com universidades internacionais e intervenientes industriais selecionados. O projeto procura, assim, desenvolver atividades de reforço das capacidades que os educadores e os responsáveis pela elaboração dos programas de estudos académicos possam utilizar para incluir na formação de futuros profissionais da IA (por exemplo, cursos sobre ética no domínio da IA e inovação responsável) considerações sobre os riscos de desvio e utilização indevida da investigação e inovação civil no domínio da IA por parte de intervenientes irresponsáveis, ligando-os a um contexto mais vasto de paz e segurança. Através destes seminários, o projeto procurará também identificar uma rede de educadores, criadores de currículos e estudantes interessados, que apoiaria a divulgação e a promoção do material educativo e das atividades de reforço das capacidades do projeto na comunidade da educação para a IA e na comunidade dos profissionais da IA. Esta componente de ligação em rede procura assegurar a sustentabilidade dos projetos para além da sua duração imediata e permitir o estabelecimento de laços mais fortes em apoio do empenhamento técnico civil em prol de objetivos mais vastos de paz, segurança, desarmamento e controlo de armas.

### 3.2.2. Descrição do projeto

Este projeto organizará uma série de seminários educativos e de reforço das capacidades com educadores e estudantes de universidades selecionadas de todo o mundo. Estes consistiriam numa combinação de palestras e atividades interativas, proporcionando aos educadores e estudantes oportunidades para refletirem sobre a forma como a investigação e a inovação no domínio civil no domínio da IA podem ser desviadas e indevidamente utilizadas de formas que possam representar riscos para a paz e a segurança internacionais e de que modo esses riscos podem ser identificados, evitados ou atenuados no processo de investigação e inovação, ou através de outros processos de governação. Estas atividades basear-se-ão no anterior trabalho piloto de menor escala realizado pelo GNUAD, que experimentou métodos para dialogar com os estudantes CTEM e sensibilizá-los para a importância de ter em conta o impacto mais vasto do seu trabalho, bem como de colaborar com conhecimentos especializados fora dos seus domínios de origem. Concretamente, estas consistiriam em

- (a) seminários regionais de reforço das capacidades para educadores e estudantes (4) – os seminários regionais realizarão e promoverão atividades que os educadores podem utilizar para reforçar a capacidade dos estudantes CTEM em matéria de inovação responsável no domínio da IA, com especial destaque para a forma de avaliar e atenuar os riscos de desvio e utilização indevida de tecnologias da IA civis por intervenientes irresponsáveis. Cada seminário será organizado com uma universidade sediada na UE e uma universidade prestigiada de uma região diferente do mundo, ligando sempre um conjunto diversificado de participantes sediados na UE a um conjunto diversificado de pessoas sediadas fora da UE. Os seminários abrangerão assim a América Latina e as Caraíbas, a América do Norte, a África e a Ásia e o Pacífico, o que permitirá a participação de estudantes (a nível de mestrado e doutoramento) de todo o mundo, incluindo do hemisfério sul. Os seminários serão realizados principalmente em inglês, mas, sempre que possível, os participantes terão a oportunidade de participar em atividades baseadas em agrupamentos linguísticos alternativos; e
- (b) num seminário internacional sobre o reforço das capacidades sustentáveis (1) – o seminário basear-se-á nos ensinamentos retirados dos seminários regionais e facilitará o intercâmbio de informações e experiências entre educadores e estudantes selecionados das universidades envolvidas no projeto. O seminário debaterá a forma de aperfeiçoar as atividades e os instrumentos elaborados ao longo do projeto e de os divulgar para além do grupo de universidades participantes. Debaterá também a forma de apoiar a participação dos estudantes numa IA responsável que aborde os riscos de desvio e de utilização indevida para a paz e a segurança internacionais após a sua entrada no mercado de trabalho.

As redes e a presença do SIPRI e do GNUAD em África, na Ásia-Pacífico, na Europa, na América do Norte e do Sul serão utilizadas para facilitar e apoiar aspetos das atividades, conforme adequado.

### **3.2.3. Resultados esperados do projeto**

Espera-se que o projeto crie modelos de desenvolvimento de capacidades e atividades de participação que os educadores e os responsáveis pela elaboração dos programas de estudos académicos possam replicar para sensibilizar os futuros profissionais da IA para os problemas do desvio e da utilização indevida da IA civil por intervenientes irresponsáveis e para a forma como podem ajudar a atenuar estes problemas através de processos de inovação responsáveis. Após a conclusão das atividades do projeto, espera-se que os participantes (educadores, mas também estudantes CTEM) possam utilizar e promover ferramentas, métodos e conceitos de inovação responsáveis para identificar e atenuar os riscos de desvio e utilização indevida no desenvolvimento e difusão da investigação e inovação civil no domínio da IA.

Espera-se igualmente que as atividades do projeto criem uma rede de educadores, responsáveis pela elaboração dos programas de estudos académicos e estudantes que não só promovam as atividades do projeto no âmbito das comunidades de educação e profissionais no domínio da IA (por exemplo, durante as conferências da *Computational Intelligence Society* do IEEE), mas também estejam disponíveis para contribuir com capacidade técnica para processos de governação internacional com liderança estatal (por exemplo, o processo da Convenção sobre Certas Armas Convencionais relativo a tecnologias emergentes no domínio dos sistemas de armas letais autónomos).

O valor destas atividades a curto e a longo prazo será demonstrado através de inquéritos pré e pós-atividade.

### **3.3. Projeto 3 – Facilitar o desenvolvimento sustentável, a divulgação e o impacto a longo prazo da inovação responsável na IA para a paz e a segurança**

#### **3.3.1. Objetivo do projeto**

O objetivo do projeto 3 é facilitar o desenvolvimento sustentável a longo prazo, a divulgação e o impacto da inovação responsável no domínio da IA, como forma de atenuar os riscos que o desvio e a utilização indevida da investigação e inovação civil no domínio da IA podem representar para a paz e a segurança. Pretende fazê-lo através de mesas-redondas com a indústria da IA, de diálogos multilaterais, da criação de um relatório público e de atividades de divulgação específicas. O projeto visa assegurar que o trabalho gerado, em particular as atividades de educação, reforço de capacidades e participação, alcança e influencia a comunidade da IA em geral, a todos os níveis (desde os estudantes aos engenheiros e outros profissionais da IA) e entre fronteiras geográficas, setoriais e outras. Para aumentar a possibilidade de ter um impacto amplo e profundo, é essencial cooperar com organizações profissionais neste domínio, como o IEEE, e realizar compromissos multidimensionais em todo o meio académico, na indústria e noutros setores. Estes esforços darão aos representantes interessados de diferentes comunidades da IA a oportunidade de se apropriarem do problema e de apresentarem os seus próprios pontos de vista sobre a forma como os esforços de atenuação dos riscos podem ser realizados e promovidos de forma sustentável nas respetivas comunidades e em toda a comunidade mundial da IA. É igualmente importante, para o valor a longo prazo do projeto para os Estados, as organizações intergovernamentais e outros, que os profissionais da IA possam aprender com peritos governamentais envolvidos na atenuação dos riscos no contexto do desarmamento, do controlo de armas e da não proliferação e colaborar com eles. É também fundamental para a sustentabilidade assegurar que as perspetivas geradas pelas atividades de envolvimento são analisadas, consolidadas e divulgadas de forma adequada.

### 3.3.2. Descrição do projeto

Este projeto é composto pelas seguintes vertentes fundamentais:

- (a) diálogos multilaterais sobre "inovação responsável no domínio da IA para a paz e a segurança" (até 9) – esta série de reuniões virtuais de diálogo juntará peritos do meio académico, da investigação, do setor privado e do controlo tradicional do armamento da UE e de outros países para debater
- i. tendências tecnológicas que possam gerar riscos de difusão, de desvio e utilização indevida, com impactos para a paz e a segurança internacionais;
  - ii. como participar na atenuação dos riscos através de processos, métodos e meios de inovação responsáveis, bem como oportunidades e desafios para o diálogo e a partilha de conhecimentos entre as comunidades de partes interessadas, incluindo as que operam noutros setores, como as indústrias biológicas e químicas; e
  - iii. o valor potencial, a finalidade e o formato de uma rede autossustentável de peritos e atividades de diálogo. O grupo de peritos reunir-se-á várias vezes por ano e trabalhará no sentido de organizar dois eventos públicos para a comunidade em geral.

Das nove reuniões virtuais, duas deverão ser abertas ao público, a fim de facilitar uma consulta mais ampla.

- (b) mesas-redondas do setor privado (até 6) – esta série de mesas-redondas virtuais dará início a um diálogo com os intervenientes que trabalham com processos de inovação responsáveis no domínio da IA no setor privado (por exemplo, a parceria para a inteligência artificial) sobre a forma como podem contribuir para minimizar os riscos de desvio e utilização indevida de tecnologias de IA civis por intervenientes irresponsáveis, bem como explorar possíveis incentivos no âmbito do desenvolvimento do setor privado para o efeito. Os temas incluirão:
- i. a relevância do contexto internacional de segurança e desarmamento para o setor privado;
  - ii. o(s) quadro(s) jurídico(s) em que o desenvolvimento, a implantação e o funcionamento da IA existem em todo o mundo;
  - iii. como tirar partido dos mecanismos de avaliação dos riscos e de outras medidas que fazem parte ou podem ser integradas em processos de inovação responsáveis e em programas de conformidade das empresas; e
  - iv. ensinamentos a retirar de outros setores, processos e quadros relacionados com o controlo de armas (por exemplo, indústrias biológicas e químicas).
- (c) relatório sobre as perspetivas da comunidade da IA sobre o controlo de armas e a atenuação dos riscos na IA, orientado para as comunidades da IA e do controlo de armas (1) – a elaboração deste relatório irá captar e consolidar as conclusões e recomendações do projeto num documento de referência único destinado tanto às comunidades da IA responsável civil como a do controlo de armas. O relatório debaterá a forma como os riscos para a paz e a segurança internacionais associados ao desvio e à utilização indevida da investigação e inovação no domínio da IA civil podem ser identificados, avaliados e abordados.

(d) eventos de divulgação orientados para a consulta e o envolvimento com as comunidades da IA e do controlo de armas (a determinar) – a equipa procurará oportunidades para comunicar o trabalho e as suas conclusões e recomendações ao longo de todo o projeto. O formato dos eventos e o conteúdo das apresentações serão adaptados às necessidades dos grupos-alvo. Estes podem incluir reuniões do Grupo CONOP, a assembleia da Aliança Europeia da IA; o Grupo de Peritos Governamentais sobre tecnologias emergentes no domínio dos sistemas de armas letais autónomos; o grupo de trabalho interagências sobre a IA (IAWG-AI); a iniciativa "AI for Good", da UIT; o diálogo anual sobre inovação do UNIDIR; e o Instituto dos Engenheiros de Eletricidade e de Eletrónica. A equipa procurará também colaborar bilateralmente com as partes interessadas pertinentes de governos, do meio académico e do setor privado.

### **3.3.3. Resultados esperados do projeto**

Espera-se que este projeto crie as bases para o desenvolvimento sustentável, a divulgação e o impacto da inovação responsável dos processos da IA, abordando os riscos de difusão, desvio e utilização indevida e as suas implicações para a paz e a segurança para além da duração imediata da decisão do Conselho.

O diálogo multilateral deverá proporcionar um modelo para a partilha de informações e a colaboração em matéria de atenuação dos riscos, não só no seio da comunidade mundial da IA, mas também entre a comunidade da IA responsável civil e as comunidades do desarmamento, do controlo de armas e da não proliferação. Esse modelo poderia ser utilizado para familiarizar os decisores políticos com os principais avanços tecnológicos e científicos relevantes para a inovação responsável da IA, bem como para familiarizar os públicos técnicos com o ambiente em que os decisores políticos estão atualmente envolvidos. Espera-se que o projeto facilite relações sustentáveis e o envolvimento entre os intervenientes interessados dentro e entre estas diferentes comunidades. Espera-se que esses efeitos de rede de contactos heterogéneos permitam um maior desenvolvimento e uma promoção generalizada da inovação responsável da IA para a paz e a segurança para além do calendário do projeto.

O diálogo com o setor privado deverá permitir um maior e mais profundo envolvimento do setor privado da IA na identificação, prevenção e atenuação dos riscos para a paz e a segurança decorrentes do desvio e da utilização indevida da investigação e inovação no domínio da IA civil. Espera-se que o projeto proporcione aos principais intervenientes nos processos do setor privado uma maior compreensão e apropriação dos problemas que procura resolver. Além disso, visa facilitar a (mais ampla) adoção e execução de processos, métodos e meios de inovação responsáveis nos atuais mecanismos e procedimentos de gestão dos riscos empresariais.

O diálogo multilateral e a mesa-redonda do setor privado deverão também gerar informações sobre uma série de questões substanciais, incluindo a) a forma como os métodos e meios de inovação responsáveis podem ser aperfeiçoados e utilizados para identificar, prevenir e atenuar os riscos decorrentes do desvio e da utilização abusiva da investigação e inovação civil no domínio da IA; b) a forma como a investigação e a inovação no domínio da IA podem ser utilizadas de forma positiva para apoiar os objetivos de paz e segurança (por exemplo, aplicações de alerta precoce em caso de conflito e de assistência humanitária); e c) como facilitar um maior diálogo e partilha de informações entre os esforços de atenuação dos riscos empreendidos na comunidade da IA responsável civil (ou seja, várias iniciativas lideradas pelo IEEE) com os esforços existentes na comunidade do desarmamento, do controlo de armas e da não proliferação a nível intergovernamental.

O relatório e as atividades de divulgação permitirão analisar, consolidar e divulgar os pontos de vista gerados através dos projetos 1, 2 e 3, apoiando assim a promoção das conclusões e recomendações das atividades do projeto na comunidade mundial da IA, bem como na comunidade política. Espera-se igualmente que contribuam para assegurar a sustentabilidade do impacto para além do calendário dos projetos.

#### **4. DURAÇÃO**

A duração total estimada de execução dos projetos é de 36 meses.