



Consejo de la  
Unión Europea

Bruselas, 26 de octubre de 2022  
(OR. en)

14095/22  
ADD 1

LIMITE

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

## PROPUESTA

---

De:	Por el alto representante de la Unión para Asuntos Exteriores y Política de Seguridad, D. Stefano SANNINO, secretario general
Fecha de recepción:	26 de octubre de 2022
A:	D. <sup>a</sup> Thérèse BLANCHET, secretaria general del Consejo de la Unión Europea
Asunto:	Propuesta del Alto Representante de la Unión para Asuntos Exteriores y Política de Seguridad al Consejo de Decisión del Consejo en apoyo de la ejecución del proyecto «Promover la innovación responsable en inteligencia artificial para la paz y la seguridad»

---

Adjunto se remite a las Delegaciones el documento – HR(2022) 238.

Adj.: HR(2022) 238

# SERVICIO EUROPEO DE ACCIÓN EXTERIOR



## DOCUMENTO DE PROYECTO

### **Promover la innovación responsable en inteligencia artificial para la paz y la seguridad**

**HR(2022)238**

#### **1. ANTECEDENTES**

Los recientes avances en inteligencia artificial (IA) han abierto nuevas posibilidades para reforzar y mantener la paz y la seguridad, por ejemplo, a través de mejoras tecnológicas en ámbitos como la alerta rápida de conflictos y la verificación del control de armamento y exportaciones. Por otra parte, estos avances han proporcionado nuevos medios para generar o agravar tensiones, conflictos e inseguridad entre los Estados y dentro de ellos. Los riesgos que plantean determinadas aplicaciones de la IA, como los sistemas de armas autónomos letales, son un motivo de gran preocupación para la comunidad de control de armamento. Una vía de riesgo que merece una mayor atención, y frente a la que los actuales esfuerzos diplomáticos y de control de armamento pueden ser insuficientes para ofrecer una respuesta, es el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables (en particular agentes no estatales malintencionados) como por ejemplo el uso indebido de redes generativas antagónicas para producir «ultrafalsificaciones» para campañas de desinformación.

La IA es una tecnología facilitadora con un gran potencial de uso general. La investigación y la innovación en materia de IA desarrolladas para usos civiles podrían ser accesibles y readaptadas (con relativa facilidad) por determinados agentes para usos nocivos o disruptivos que podrían tener implicaciones para la paz y la seguridad internacionales. El desvío y el uso indebido de la tecnología civil no son fenómenos nuevos ni exclusivos de la IA. Un ejemplo claro y reciente lo tenemos en el ámbito afín de la robótica, con la militarización y el uso de drones recreativos por parte del Daesh/ISIS en Siria. Sin embargo, en el caso de la IA, el problema se complica por múltiples factores: la naturaleza intangible y de rápida evolución de los algoritmos y datos de la IA, lo que dificulta el control de su transferencia y proliferación; el liderazgo del sector privado en el mundo de la investigación, desarrollo e innovación, y su consiguiente necesidad de proteger los algoritmos de su propiedad; y la disponibilidad mundial de conocimientos especializados humanos y recursos materiales capaces de reconvertir las tecnologías de IA. Mientras tanto, quienes trabajan con IA en el sector civil demasiado a menudo no son

conscientes de las posibles implicaciones que el desvío y el uso indebido de su trabajo podrían tener para la paz y la seguridad internacionales, o se muestran reticentes a participar en los debates existentes en torno a los riesgos de la IA en los ámbitos del control de armamento y la no proliferación.

Es necesario respaldar una mayor implicación de la comunidad civil de la IA en la comprensión y reducción de los riesgos para la paz y la seguridad asociados al desvío y el uso indebido de la tecnología civil de IA por parte de agentes irresponsables. Según el Instituto Internacional de Estocolmo de Investigaciones para la Paz (SIPRI) y la Oficina de Asuntos de Desarme de las Naciones Unidas (UNODA), esto podría lograrse mediante un mayor fomento de la innovación responsable en la comunidad civil mundial de la IA. El trabajo llevado a cabo en el pasado por el SIPRI y la UNODA ha demostrado que la innovación responsable, como mecanismo de autogobierno, podría proporcionar a la comunidad civil mundial de la IA herramientas y métodos prácticos para identificar y ayudar a prevenir y reducir los riesgos que el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA podrían plantear para la paz y la seguridad. El trabajo del SIPRI y la UNODA también identificó metodologías y varias iniciativas civiles en curso sobre IA responsable que podrían aprovecharse para sensibilizar a la comunidad civil de la IA sobre cuestiones relacionadas con el control de armamento y la no proliferación, debates de expertos y posiciones estatales sobre el desarrollo responsable, la difusión y el uso de la IA, así como las lecciones que deben extraerse del trabajo en relación con la responsabilidad en el sector de la defensa<sup>1</sup>. Lo más importante es que este trabajo previo ha identificado claramente la participación de estudiantes de ciencia, tecnología, ingeniería y matemáticas (CTIM), que participan en la IA desde un planteamiento educativo, como elemento central de cualquier esfuerzo eficaz de innovación responsable.

## **2. OBJETIVOS**

Estos proyectos tienen por objeto respaldar una mayor participación de la comunidad civil de la IA en la reducción de los riesgos que el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables plantean para la paz y la seguridad internacionales. Para ello tienen la intención, en primer lugar, de generar una mayor comprensión de cómo las decisiones en el desarrollo y la difusión de la investigación y la innovación en materia de IA pueden afectar a los riesgos de desvío y uso indebido, y a su vez suponer riesgos u oportunidades para la paz y la seguridad, y, en segundo lugar, de promover procesos, métodos y herramientas de innovación responsables que puedan contribuir a asegurar la aplicación pacífica de las innovaciones civiles y la difusión responsable de los conocimientos sobre IA. A tal fin, respaldan actividades de capacitación, investigación y participación que *i)* mejorarán la capacidad dentro de la comunidad civil mundial de la IA para incluir y abordar, mediante procesos de innovación responsable, los riesgos para la paz y la seguridad que plantean

---

<sup>1</sup> Las metodologías incluyen, por ejemplo, las prácticas recomendadas del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) para evaluar la repercusión de los sistemas autónomos e inteligentes en el bienestar humano (IEEE Std 7010-2020) y la lista de evaluación para una IA fiable del grupo de expertos de alto nivel sobre la IA. Las iniciativas incluyen: la Iniciativa Mundial sobre la Ética de los Sistemas Inteligentes y Autónomos del IEEE; la Asociación de IA; la Asociación Mundial de IA.

el desvío y el uso indebido de la IA civil por parte de agentes irresponsables; y *ii*) reforzarán el vínculo entre los esfuerzos de reducción de riesgos en materia de IA responsable en el ámbito civil y los que ya están en curso en la comunidad de desarme, control de armamento y no proliferación a escala intergubernamental. Resulta fundamental señalar que no tienen por objeto fijar nuevas normas o principios, ni entrar en ámbitos que sean competencia de los Estados, sino que pretenden desarrollar esfuerzos civiles de innovación responsable para incluir los riesgos que el desvío y el uso indebido de la IA civil por parte de agentes irresponsables plantean para la paz y la seguridad, y educar sobre los esfuerzos intergubernamentales pertinentes en curso.

Para llegar de manera efectiva a la comunidad civil de la IA e influir en ella, los proyectos aplican un planteamiento triple, con el objetivo de:

- (a) *colaborar con los educadores*. Trabajar con determinados educadores y desarrolladores de planes de estudios académicos en el desarrollo y la promoción de materiales educativos que pueden utilizarse para integrar la consideración de los riesgos de paz y seguridad que derivan del desvío y el uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables en la formación de futuros profesionales de la IA (por ejemplo, en cursos sobre ética en la IA e innovación responsable);
- (b) *colaborar con los estudiantes*. Enseñar a una serie de estudiantes de CTIM de todo el mundo cómo identificar, prevenir o reducir los riesgos para la paz y la seguridad que plantean el desvío y el uso indebido del desarrollo civil de la IA por parte de agentes irresponsables en el proceso de investigación e innovación o a través de otros procesos de gobernanza; y
- (c) *colaborar con el sector de la IA*. Trabajar con asociaciones profesionales y organismos de normalización como el IEEE para i) difundir materiales educativos adaptados y actividades de participación entre los profesionales técnicos; ii) respaldar los usos positivos de la IA para la paz y la seguridad; y iii) facilitar el diálogo y el intercambio de información entre expertos del mundo académico, el sector privado y el sector público sobre la manera de reducir el riesgo de desvío y uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables.

Este enfoque permite que los proyectos lleguen a la comunidad de la IA a todos los niveles, incluidos no solo los profesionales actuales, sino también las generaciones futuras. Asimismo, permite la colaboración entre el mundo académico, las empresas y otros sectores, y respalda la sostenibilidad de los esfuerzos futuros mediante el establecimiento de redes que trasciendan estas fronteras.

Los proyectos también pretenden utilizar el poder de convocatoria y la experiencia del SIPRI y la UNODA para llegar a la comunidad de la IA a escala mundial, no solo a las partes interesadas de la UE. El SIPRI y la UNODA están en una posición única para llegar a los agentes de IA de África, Asia y el Pacífico, Europa y América del Norte y del Sur y facilitar la colaboración entre ellos. Ambas entidades cuentan, además, con experiencia en otros ámbitos de la ciencia y la tecnología que se enfrentan a retos similares en relación con el doble uso y la proliferación, como por ejemplo la biotecnología. Del mismo modo, los proyectos tratan de aprovechar las

condiciones existentes en la Unión Europea, como a) la existencia de procesos avanzados multilaterales sobre la IA responsable; b) el alto nivel de participación en cuestiones de desarme, control de armamento y no proliferación en la UE, y de conocimientos especializados al respecto; c) la diversidad de conexiones que las instituciones académicas, de investigación y del sector privado de la UE tienen con otras regiones, en particular en el Sur Global, que también serán un objetivo importante para la participación; y d) la diversidad de nacionalidades de estudiantes, educadores e ingenieros en universidades, centros de investigación y en el sector privado.

La inclusión será una consideración fundamental para la realización de las actividades de los proyectos. Con el fin de respaldar de manera eficaz a la comunidad de la IA, los proyectos reconocen que la comunidad de la IA está compuesta por una amplia gama de agentes y, en particular, que:

- (a) el género es un factor muy importante. Por ello, la perspectiva de género se integrará en consonancia con las estrategias de igualdad e integración de la perspectiva de género de todo el sistema de las Naciones Unidas. Se fomentará y exigirá la participación de las mujeres en todas las actividades del proyecto; y
- (b) la inclusión de las personas con discapacidad y los ajustes razonables de las necesidades se llevarán a cabo en todo el proyecto. Esto engloba abordar los obstáculos a la participación de las personas con discapacidad, así como asegurar que se adopten medidas para colaborar con las personas con discapacidad y facilitar la representación de sus ideas y experiencias.

### **3. PROYECTOS**

Los tres proyectos que se describen a continuación pretenden complementarse y respaldarse mutuamente, con elementos que se desarrollan a lo largo de los 36 meses.

#### **Proyecto 1 – Producción de material de formación y capacitación para la comunidad civil de la IA**

##### **3.1. Objetivo del proyecto**

El proyecto 1 se centra en ofrecer formación y recursos a los agentes de la comunidad civil de la IA para valorar y reducir los riesgos que puedan plantear para la paz y la seguridad internacionales el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables. Tiene por objeto producir material de formación y capacitación que proporcione a los profesionales de la IA de todas las regiones, niveles y sectores (incluidos educadores centrados en la IA, desarrolladores de planes de estudios, estudiantes de CTIM e ingenieros e investigadores de IA del mundo académico y del sector privado) la información y las herramientas necesarias para:

- (a) comprender cómo la investigación y la innovación civiles en materia de IA podrían desviarse y usarse indebidamente de formas que pudieran suponer riesgos para la paz y la seguridad internacionales, y cómo las decisiones en el desarrollo y la difusión de la investigación y la innovación aumentan o reducen el riesgo de desvío y uso indebido;

- (b) comprender los esfuerzos ya realizados por la comunidad de desarme, control de armamento y no proliferación para reducir los riesgos de desvío y uso indebido de la investigación y la innovación civiles; y
- (c) aplicar la innovación responsable de manera que reduzca el riesgo de desvío y uso indebido en el desarrollo y la difusión de la investigación y la innovación.

### 3.1.1. Descripción del proyecto

Este proyecto producirá tres conjuntos separados de materiales de formación y capacitación:

- (a) *Manual (1)*. El manual recopilará conocimientos y medios básicos para que los agentes de IA evalúen y reduzcan, en el proceso de investigación e innovación, los riesgos de desvío y uso indebido de la tecnología civil de IA por parte de agentes irresponsables. En él se verá por qué y cómo las decisiones en el desarrollo y la difusión de la investigación y la innovación afectan a los riesgos de desvío y uso indebido, y a su vez suponer riesgos u oportunidades para la paz y la seguridad; incluirá también las obligaciones pertinentes contempladas por el Derecho internacional y de control de las exportaciones, además de las consideraciones de seguridad y protección que se están debatiendo en los círculos militares, así como en materia de desarme, control de armamento y no proliferación. Asimismo, presentará ejemplos de procesos y herramientas para practicar la innovación responsable, como metodologías de evaluación de la repercusión tecnológica y modelos de evaluación del riesgo.
- (b) *Serie de pódcast (~ 10)*. Estos pódcast actuarán como un medio accesible y atractivo para que los agentes de la IA aprendan por qué y cómo los procesos de innovación responsable en materia de IA pueden reforzar la paz y la seguridad internacionales mediante la reducción de los riesgos que plantea el desvío y el uso indebido por parte de agentes irresponsables. La serie examinará temas importantes (p. ej., el patrón de desvío y uso indebido de la investigación y la innovación de doble uso o de uso general; los retos humanitarios, estratégicos y políticos asociados al posible uso indebido de la investigación e innovación civiles en materia de IA; los retos a los que se enfrentan los círculos de desarme, control de armamento y no proliferación en los esfuerzos de reducción de riesgos; cómo hacer innovación responsable a través de la evaluación de riesgos; cumplimiento del control de las exportaciones; reducción de riesgos desde el diseño; publicación responsable; conocimiento de los clientes; y experiencia en ejercicios de simulación teórica) y se estructurará en torno a entrevistas que el equipo del proyecto llevará a cabo con representantes de las comunidades pertinentes.
- (c) *Blog (9-10)*. El equipo desarrollará una serie de publicaciones de blog seleccionadas con el fin de aumentar la visibilidad de los esfuerzos que intentan trascender los límites entre las comunidades de «IA responsable» centrada en el ámbito civil y las comunidades de control de armamento y no proliferación. El blog ofrecerá una plataforma para difundir información, ideas y soluciones sobre la identificación y el tratamiento de los riesgos asociados con el desvío y el uso indebido de la IA civil en el proceso de investigación e innovación. Además, tratará de representar la diversidad de ideas y perspectivas presentes en el sector de la IA.

Estos materiales se difundirán públicamente a través de los sitios web de los agentes encargados, sus redes sociales y mediante la comunicación directa con las entidades académicas, las asociaciones profesionales civiles de IA y otros grupos adecuados pertinentes.

### **3.1.2. Resultados previstos del proyecto**

Se espera que este proyecto establezca una nueva serie de materiales que permita sensibilizar a los profesionales civiles de la IA sobre a) cómo la investigación y la innovación civiles en materia de IA pueden desviarse y usarse indebidamente de formas que planteen riesgos para la paz y la seguridad internacionales, b) cómo se abordan dichos riesgos en la comunidad de desarme, control de armamento y no proliferación, y c) cómo los profesionales de la IA pueden contribuir aún más a la reducción de dichos riesgos a través de procesos de innovación responsable.

Se espera que esto aumente la participación del sector de la IA civil en la reducción de los riesgos que el desvío y el uso indebido de la IA civil plantean para la paz y la seguridad internacionales, además de mejorar la capacidad de los profesionales técnicos para participar en los procesos pertinentes de la comunidad de desarme, control de armamento y no proliferación, así como respaldar la participación de nuevos públicos no incluidos tradicionalmente en los esfuerzos educativos en materia de desarme y no proliferación.

Asimismo, se espera que el material facilite la ejecución de los demás proyectos y sirva de base para las actividades de formación y capacitación del proyecto 2, así como para las actividades de diálogo y participación en el proyecto 3. A su vez, se espera que dichas actividades contribuyan a la producción y al perfeccionamiento del material. Se espera que este enfoque iterativo ayude a abordar los posibles obstáculos a su promoción, difusión y uso dentro de la comunidad de la IA, como por ejemplo cuestiones relacionadas con el idioma, el contenido, el contexto y la disponibilidad, que podrían limitar su repercusión a escala mundial, en particular en el Sur Global.

## **3.2. Proyecto 2 – Actividades de formación y capacitación para futuros profesionales de la IA**

### **3.2.1. Objetivo del proyecto**

El objetivo del proyecto 2 consiste en respaldar la integración del problema del desvío y el uso indebido de la investigación civil en materia de IA por parte de agentes irresponsables en la formación de las generaciones futuras de profesionales de la IA. A largo plazo, esto garantizará que los estudiantes de CTIM que configuran el futuro de la IA sean conscientes tanto de los efectos negativos que el desvío y el uso indebido de su trabajo por parte de agentes irresponsables podrían tener en la paz y la seguridad internacionales como de que dispondrán de las herramientas básicas necesarias para detectar y reducir este riesgo en el proceso de investigación e innovación.

Este proyecto consistirá en una serie de talleres de formación y capacitación con educadores y estudiantes, en colaboración con determinadas universidades internacionales y otros agentes del

sector. Así, el proyecto pretende desarrollar actividades de capacitación que los educadores y los desarrolladores de planes de estudios académicos podrían emplear para incluir en la formación de los futuros profesionales de la IA (por ejemplo, cursos sobre ética de la IA e innovación responsable) consideraciones sobre los riesgos de desvío y uso indebido de la investigación y la innovación civiles en materia de IA por parte de agentes irresponsables, y relacionarlos con el contexto más amplio de la paz y la seguridad. A través de estos talleres, el proyecto también intentará identificar una red de educadores, desarrolladores de planes de estudios y estudiantes interesados que respaldarían la difusión y la promoción del material educativo del proyecto y las actividades de desarrollo de capacidades en la comunidad educativa en materia de IA y en la comunidad de profesionales de la IA. Este componente de creación de redes tiene por objeto garantizar la sostenibilidad de los proyectos más allá de su duración inmediata y permitir la creación de vínculos más sólidos que sustenten el compromiso técnico civil en pro de objetivos más amplios de paz, seguridad, desarme y control de armamento.

### **3.2.2. Descripción del proyecto**

Este proyecto consistirá en una serie de talleres de formación y capacitación con educadores y estudiantes, procedentes de universidades seleccionadas de todo el mundo. Los talleres consistirán en una combinación de conferencias y actividades interactivas que ofrecerán a los educadores y a los estudiantes la posibilidad de reflexionar sobre cómo la investigación y la innovación civiles en materia de IA podrían desviarse y utilizarse indebidamente de formas que suponen riesgos para la paz y la seguridad internacionales, y cómo identificar, evitar o reducir dichos riesgos en el proceso de investigación e innovación, o a través de otros procesos de gobernanza. Estas actividades se basarán en proyectos piloto previos de menor escala llevados a cabo por la UNODA, en los que se probaron métodos para colaborar con los estudiantes de CTIM y sensibilizarlos sobre la importancia de tener en cuenta la repercusión más amplia de su trabajo, así como de colaborar con conocimientos especializados fuera de sus campos habituales. En concreto, los talleres consistirán en:

- (a) talleres regionales de capacitación para educadores y estudiantes (4). Estos llevarán a cabo y promoverán actividades que los educadores pueden utilizar para formar a los estudiantes de CTIM sobre la innovación responsable en materia de IA, prestando especial atención a la manera de evaluar y reducir los riesgos de desvío y uso indebido de la tecnología civil de IA por parte de agentes irresponsables. Cada taller se organizará con una universidad con sede en la UE y una universidad de alto nivel de una región mundial distinta, conectando así siempre a un conjunto diverso de participantes con sede en la UE con otro conjunto diverso con sede fuera de la UE. Los talleres abarcarán América Latina y el Caribe, América del Norte, África y Asia y el Pacífico, lo que permitirá la participación de estudiantes (a nivel de máster y doctorado) de todo el mundo, también del Sur Global. Los talleres se desarrollarán principalmente en inglés, pero, cuando sea posible, se permitirá a los participantes asistir a actividades basadas en agrupaciones lingüísticas alternativas; y
- (b) taller internacional sobre capacitación sostenible (1). Este se basará en las experiencias extraídas de los talleres regionales y facilitará el intercambio de información y experiencias entre educadores y algunos estudiantes seleccionados de las universidades

participantes en el proyecto. En el taller se debatirá cómo perfeccionar las actividades y herramientas elaboradas a lo largo del proyecto y difundirlas más allá del grupo de universidades participantes, además de cómo respaldar la colaboración de los estudiantes, una vez que estén en el mercado laboral, en una IA responsable que aborde los riesgos de desvío y uso indebido para la paz y la seguridad internacionales.

Las redes del SIPRI y la UNODA y la presencia de participantes de África, Asia-Pacífico, Europa, América del Norte y América del Sur se utilizarán para organizar aspectos de las actividades, según proceda, y para apoyarlos.

### **3.2.3. Resultados previstos del proyecto**

Se espera que este proyecto cree modelos de capacitación y actividades de participación que los educadores y desarrolladores de planes de estudios académicos puedan reproducir para sensibilizar a futuros profesionales de la IA sobre los problemas del desvío y el uso indebido de la IA civil por parte de agentes irresponsables y sobre cómo pueden ayudar a reducir estos problemas a través de procesos de innovación responsable. Una vez finalizadas las actividades del proyecto, se espera que los participantes (educadores, pero también estudiantes de CTIM) utilicen y promuevan herramientas, métodos y conceptos de innovación responsable para identificar y reducir los riesgos de desvío y uso indebido en el desarrollo y la difusión de la investigación y la innovación civiles en materia de IA.

Asimismo, se espera que las actividades del proyecto generen una red de educadores, desarrolladores de planes de estudios y estudiantes que no solo promuevan las actividades del proyecto dentro de las comunidades de formación sobre IA y las comunidades profesionales (por ejemplo, durante las conferencias de la Sociedad de Inteligencia Computacional del IEEE), sino que también estén disponibles para contribuir con capacidad técnica a los procesos de gobernanza internacional dirigidos por los Estados (por ejemplo, el proceso de la Convención sobre ciertas armas convencionales sobre tecnologías emergentes en el ámbito de los sistemas de armas autónomos letales).

El valor a corto y largo plazo de estas actividades se demostrará mediante encuestas previas y posteriores a la actividad.

## **3.3. Proyecto 3 – Facilitar el desarrollo, la difusión y la repercusión sostenibles y a más largo plazo de la innovación responsable en IA para la paz y la seguridad**

### **3.3.1. Objetivo del proyecto**

El objetivo del proyecto 3 es facilitar el desarrollo, la difusión y la repercusión sostenibles y a más largo plazo de la innovación responsable en IA como medio para reducir los riesgos que el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA pueden plantear para la paz y la seguridad. Para ello se celebrarán mesas redondas con el sector de la IA y diálogos multilaterales y se desarrollará un informe público y actividades específicas de divulgación de la información. El proyecto tiene por objeto asegurar que el trabajo generado, en

particular las actividades de formación, capacitación y participación, llegue a la comunidad de la IA en general —a todos los niveles (desde los estudiantes hasta los ingenieros y otros profesionales de la IA) y a través de las fronteras geográficas, sectoriales y de otro tipo— y tenga una repercusión para ella. Para aumentar las posibilidades de conseguir una repercusión amplia y profunda resulta primordial cooperar con las organizaciones profesionales de este sector, como el IEEE, y realizar actividades multidimensionales en el mundo académico, el sector empresarial y otros sectores. Estos esfuerzos permitirán a los representantes interesados de diferentes comunidades de la IA asumir la responsabilidad del problema y aportar sus propios puntos de vista sobre cómo llevar a cabo los esfuerzos de reducción de riesgos y promoverlos de manera sostenible dentro de la comunidad mundial de la IA y entre sus diferentes partes. Asimismo, en lo que se refiere al valor a largo plazo que el proyecto tiene para los Estados, las organizaciones intergubernamentales y otros agentes, es importante que los profesionales de la IA, por una parte, aprendan de expertos gubernamentales que participan en la reducción de riesgos en el contexto del desarme, el control de armamento y la no proliferación y, por otra parte, que colaboren con ellos. También es fundamental para la sostenibilidad asegurar el análisis, la consolidación y la divulgación adecuados de los conocimientos generados a través de las actividades de participación.

### **3.3.2. Descripción del proyecto**

Este proyecto consta de las siguientes líneas principales:

- (a) diálogos multilaterales sobre «una innovación responsable en materia de IA para la paz y la seguridad» (hasta nueve). Esta serie de reuniones virtuales congregará a expertos del mundo académico, el mundo de la investigación, el sector privado y el sector del control de armamento tradicional de la UE y fuera de ella para debatir:
- i. tendencias tecnológicas que puedan generar riesgos de difusión, desvío y uso indebido con repercusiones para la paz y la seguridad internacionales;
  - ii. la manera de participar en la reducción de riesgos a través de procesos, métodos y medios de innovación responsable, y oportunidades y retos para el diálogo y el intercambio de conocimientos entre las comunidades de partes interesadas, incluidas las que operan en otros sectores, como las industrias biológicas y químicas; y
  - iii. el valor potencial, la finalidad y el formato de una red autosuficiente de expertos y actividades de diálogo. El grupo de expertos se reunirá varias veces al año y trabajará en la organización de dos actos públicos para la comunidad en general.

De las nueve reuniones virtuales, dos estarán abiertas al público, con el fin de facilitar una consulta más amplia.

- (b) mesas redondas con el sector privado (hasta seis). Esta serie de mesas redondas virtuales dará pie a un diálogo con los agentes que trabajan con procesos de innovación en IA responsable en el sector privado (por ejemplo, la Asociación de IA) sobre la manera de contribuir a minimizar los riesgos de desvío y uso indebido de las tecnologías

civiles de IA por parte de agentes irresponsables, así como sobre la manera de explorar posibles incentivos del sector privado para el desarrollo. Los temas incluyen:

- i. la importancia del contexto internacional de seguridad y desarme para el sector privado;
- ii. el entorno o entornos jurídicos en los que se producen el desarrollo, el despliegue y la explotación de la IA en todo el mundo;
- iii. la manera de aprovechar los mecanismos de evaluación de riesgos y otras medidas que forman parte de los procesos de innovación responsable y de los programas de cumplimiento corporativo o que podrían integrarse en ellos; y
- iv. lecciones que deben extraerse de otros sectores, procesos y marcos relacionados con el control de armamento (por ejemplo, empresas biológicas y químicas).

(c) informe sobre las perspectivas de la comunidad de la IA en materia de control de armamento y reducción de riesgos en la IA, dirigido a las comunidades de IA y de control de armamento (uno). El desarrollo de este informe recopilará y consolidará las conclusiones y recomendaciones del proyecto en un único documento de referencia destinado tanto a las comunidades civiles de IA responsable como las de control de armamento. En el informe se debatirá cómo identificar, evaluar y abordar los riesgos internacionales para la paz y la seguridad relacionados con el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA.

(d) actos de divulgación destinados a la consulta y a la implicación con las comunidades de la IA y del control de armamento (por determinar). El equipo buscará oportunidades para transmitir el trabajo y sus conclusiones y recomendaciones a lo largo de todo el proyecto. El formato de los eventos y el contenido de las presentaciones se adaptarán a las necesidades de los grupos destinatarios: podrá tratarse de reuniones del Grupo «No Proliferación y Exportación de Armas» (CONOP), de la asamblea de la Alianza europea de la IA, del grupo de expertos gubernamentales sobre las tecnologías emergentes en el ámbito de los sistemas de armas autónomos letales, del Grupo de Trabajo Interinstitucional sobre IA (IAWG-IA), de la iniciativa de AI for Good (IA para el bien) de la UIT, del Diálogo sobre Innovación anual del UNIDIR y del IEEE. El equipo también tratará de colaborar bilateralmente con las partes interesadas pertinentes de las administraciones públicas, del mundo académico y del sector privado.

### **3.3.3. Resultados previstos del proyecto**

Se espera que este proyecto sienta las bases para el desarrollo sostenible, la difusión y la incidencia de la innovación responsable de los procesos de IA que abordan los riesgos de difusión, desvío y uso indebido y sus implicaciones para la paz y la seguridad más allá de la duración inmediata de la decisión del Consejo.

Se espera que el diálogo multilateral proporcione un modelo para el intercambio de información y la colaboración en materia de reducción de riesgos no solo dentro de la comunidad mundial de la IA, sino también entre la comunidad civil de la IA responsable y las comunidades de desarme, control de armamento y no proliferación. Dicho modelo podría utilizarse para familiarizar a los responsables políticos con los avances tecnológicos y científicos clave que resulten pertinentes

para la innovación responsable de la IA, así como para familiarizar al público técnico con el entorno en el que operan actualmente los responsables políticos. Se espera que el proyecto facilite las relaciones y la participación sostenibles entre los agentes interesados dentro de estas distintas comunidades y entre ellas. Se espera que estos efectos de contactos heterogéneos permitan un mayor desarrollo y una promoción generalizada de la innovación responsable de la IA para la paz y la seguridad más allá de la duración del proyecto.

Asimismo, se espera que el diálogo con el sector privado permita una participación mayor y más decisiva del sector privado de la IA en la identificación, la prevención y la reducción de los riesgos para la paz y la seguridad derivados del desvío y el uso indebido de la investigación y la innovación civiles en materia de IA. Se espera que el proyecto ofrezca a los agentes clave de los procesos del sector privado una mejor comprensión y asunción de la responsabilidad de los problemas que pretende resolver. Además, tiene por objeto fomentar una (mayor) adopción y aplicación de procesos, métodos y medios de innovación responsable en los mecanismos y procedimientos corporativos existentes de gestión de riesgos.

Asimismo, se espera que el diálogo multilateral y la mesa redonda del sector privado generen conocimientos sobre una serie de cuestiones importantes, entre las que se encuentran a) cómo perfeccionar y emplear métodos y medios de innovación responsable para detectar, prevenir y reducir los riesgos que plantean el desvío y el uso indebido de la investigación y la innovación civiles en materia de IA; b) cómo utilizar la investigación y la innovación en materia de IA de manera positiva para respaldar los objetivos de paz y seguridad (por ejemplo, aplicaciones de alerta rápida de conflictos y asistencia humanitaria); y c) cómo facilitar un mayor diálogo y un mayor intercambio de información entre las iniciativas de reducción de riesgos emprendidas en la comunidad civil de la IA responsable (es decir, diversas iniciativas dirigidas por el IEEE) y aquellas que ya están en curso en la comunidad de desarme, control de armamento y no proliferación a nivel intergubernamental.

El informe y las actividades de divulgación analizarán, consolidarán y transmitirán los conocimientos generados a través de los proyectos 1, 2 y 3 y, de este modo, respaldarán la promoción de las conclusiones y recomendaciones de las actividades del proyecto en la comunidad mundial de la IA, así como en la comunidad política. Asimismo, se espera que contribuyan a garantizar la sostenibilidad de su repercusión más allá de la duración de los proyectos.

#### **4. DURACIÓN**

La duración total prevista de la ejecución de los proyectos será de 36 meses.