



Conseil de
l'Union européenne

Bruxelles, le 26 octobre 2022
(OR. en)

14095/22
ADD 1

LIMITE

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

PROPOSITION

Origine:	Pour le haut représentant de l'Union pour les affaires étrangères et la politique de sécurité, Monsieur Stefano SANNINO, secrétaire général
Date de réception:	26 octobre 2022
Destinataire:	Madame Thérèse BLANCHET, secrétaire générale du Conseil de l'Union européenne
Objet:	Proposition présentée au Conseil par le haut représentant de l'Union pour les affaires étrangères et la politique de sécurité en vue d'une décision du Conseil visant à soutenir la mise en œuvre d'un projet "Promouvoir l'innovation responsable dans le domaine de l'intelligence artificielle pour favoriser la paix et la sécurité"

Les délégations trouveront ci-joint le document HR(2022) 238.

p.j.: HR(2022) 238

SERVICE EUROPÉEN POUR L'ACTION EXTÉRIEURE



DOCUMENT DE PROJET

Promouvoir l'innovation responsable dans le domaine de l'IA pour favoriser la paix et la sécurité

HR(2022)238

1. CONTEXTE

Les avancées récentes accomplies dans le domaine de l'intelligence artificielle (IA) ont ouvert de nouvelles possibilités pour ce qui est de soutenir et de maintenir la paix et la sécurité, grâce par exemple à des améliorations technologiques dans des domaines tels que l'alerte précoce en cas de conflit et la vérification en matière de contrôle des armements et des exportations. D'autre part, ces avancées ont rendu possibles de nouveaux moyens de faire naître des tensions, des conflits et une insécurité entre les États et au sein de ceux-ci, ou d'aggraver des situations existantes. Les risques posés par certaines applications de l'IA, par exemple les systèmes d'armes létales autonomes, sont devenus un sujet de préoccupation majeur pour la communauté de la maîtrise des armements. Le détournement et l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables, y compris des acteurs non étatiques malveillants, dans le cadre par exemple de l'utilisation abusive de réseaux antagonistes génératifs (GAN) pour produire des "trucages vidéo ultra-réalistes" (deepfakes) en vue de la réalisation de campagnes de désinformation, constituent un risque qui mérite une plus grande attention et face auquel les efforts déployés à l'heure actuelle en matière de maîtrise des armements et de diplomatie pourraient être insuffisants.

L'IA est une technologie générique qui offre de nombreuses possibilités d'utilisation à grande échelle. Certains acteurs pourraient accéder (relativement facilement) à la recherche et à l'innovation dans le domaine de l'IA menées pour des applications civiles et les réutiliser à des fins préjudiciables ou perturbatrices, qui pourraient avoir des répercussions sur la paix et la sécurité internationales. Le détournement et l'utilisation abusive de technologies civiles ne sont pas des phénomènes nouveaux et ne sont pas propres à l'IA. Dans le domaine connexe de la robotique, l'utilisation comme arme et l'instrumentalisation récentes de drones à usage récréatif par l'EIIL/Daech en Syrie en constituent un exemple représentatif. Mais dans le cas de l'IA, le problème est compliqué par de multiples facteurs: la nature immatérielle et en rapide évolution des algorithmes et des données de l'IA, qui en rend le transfert/la prolifération difficile à contrôler; le rôle moteur du secteur privé dans l'écosystème de la recherche, du développement et de l'innovation, et la nécessité qui en découle de protéger les algorithmes propriétaires; et la possibilité d'accéder, à l'échelle mondiale, à l'expertise humaine et aux ressources matérielles capables de réutiliser les technologies de l'IA. Dans le même temps, les personnes travaillant dans le domaine de l'IA dans le secteur civil ignorent trop souvent les implications potentielles que le détournement et l'utilisation abusive de leur travail pourraient avoir sur la paix et la sécurité internationales ou hésitent à prendre part aux discussions existantes sur les risques liés à l'IA dans les milieux de la maîtrise des armements et de la non-prolifération.

Il est nécessaire de contribuer à ce que la communauté civile de l'IA se mobilise davantage afin de comprendre et d'atténuer les risques pour la paix et la sécurité liés au détournement et à l'utilisation abusive, par des acteurs irresponsables, des technologies civiles dans le domaine de l'IA. Pour l'Institut international de recherche sur la paix de Stockholm (SIPRI) et le Bureau des affaires de désarmement des Nations unies (UNODA), cet objectif pourrait être atteint en promouvant davantage l'innovation responsable au sein de la communauté civile mondiale de l'IA. Les travaux antérieurs du SIPRI et du Bureau des affaires de désarmement des Nations unies ont montré que l'innovation responsable, en tant que mécanisme d'autogouvernance, pourrait fournir à la communauté civile mondiale de l'IA des outils et des méthodes pratiques pour recenser et contribuer à prévenir et à atténuer les risques que le détournement et l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA pourraient faire peser sur la paix et la sécurité. Les travaux du SIPRI et du Bureau des affaires de désarmement ont également permis de recenser des méthodologies et plusieurs initiatives en cours en matière d'IA responsable axées sur les efforts d'experts issus de la sphère civile qui pourraient être mises à profit pour sensibiliser la communauté civile de l'IA aux questions de maîtrise des armements et de non-prolifération, aux débats d'experts et aux positions des États sur le développement, la diffusion et l'utilisation responsables de l'IA, ainsi que des enseignements à tirer des travaux menés dans le domaine de la responsabilité du secteur de la défense¹. De manière plus fondamentale, ces travaux antérieurs ont clairement établi que la coopération avec les étudiants en sciences, technologies, ingénierie et mathématiques (STEM), qui continuent de s'investir dans l'IA dans un contexte éducatif, est au cœur de tout effort efficace d'innovation responsable.

2. OBJECTIFS

Ces projets visent à soutenir une mobilisation accrue de la communauté civile de l'IA en faveur de l'atténuation des risques que le détournement et l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables peuvent faire peser sur la paix et la sécurité internationales: Pour ce faire, ils visent, d'une part, à améliorer la compréhension de la manière dont les décisions relatives au développement et à la diffusion de la recherche et de l'innovation dans le domaine de l'IA peuvent avoir une incidence sur les risques de détournement et d'utilisation abusive et, à leur tour, faire naître des risques ou des possibilités pour la paix et la sécurité, et, d'autre part, à promouvoir des processus, méthodes et outils en matière d'innovation responsable qui peuvent contribuer à assurer l'application pacifique des innovations civiles et la diffusion responsable des connaissances relatives à l'IA. A cette fin, ils soutiennent les activités de renforcement des capacités, de recherche et de mobilisation qui i) renforceront la capacité au sein de la communauté civile mondiale de l'IA à intégrer et à traiter les risques pour la paix et la sécurité que présentent le détournement et l'utilisation abusive de l'IA civile par des acteurs irresponsables, au moyen de processus fondés sur l'innovation responsable; et ii) consolideront le lien entre les efforts d'atténuation des risques en matière d'IA responsable dans le domaine civil et ceux déjà entrepris dans la communauté du désarmement, de la maîtrise des armements et de la non-prolifération au niveau intergouvernemental. Il est essentiel de noter que les activités du projet n'ont pas pour objet d'établir de nouveaux principes, normes ou réglementations, ni d'intervenir d'une autre manière dans des domaines relevant de la compétence des États. L'intention est plutôt de développer les efforts civils en matière d'innovation responsable afin d'y inclure les risques pour la paix et la sécurité que représentent le détournement et l'utilisation abusive de l'IA civile par des acteurs irresponsables, et de proposer des activités pédagogiques sur les efforts intergouvernementaux pertinents existants en la matière.

¹ Ces méthodologies comprennent, par exemple, les pratiques recommandées par l'Institut de l'ingénierie électrique et électronique (IEEE) pour évaluer l'impact des systèmes autonomes et intelligents sur le bien-être humain (IEEE Std 7010-2020), ainsi que la liste d'évaluation pour une intelligence artificielle digne de confiance (ALTAI) établie par le groupe d'experts de haut niveau sur l'intelligence artificielle. Parmi les initiatives figurent: l'Initiative mondiale sur l'éthique des systèmes autonomes et intelligents, mise en place par l'IEEE; le partenariat sur l'IA; le partenariat mondial pour l'IA.

Pour toucher et influencer efficacement la communauté civile de l'IA, les projets déploient une approche en trois volets, visant à:

- (a) établir des liens avec les éducateurs - coopérer avec les éducateurs et concepteurs de programmes d'études universitaires choisis sur la mise au point et la promotion de matériel pédagogique pouvant être utilisé pour intégrer la prise en compte, dans la formation des futurs praticiens de l'IA (par exemple, dans des cours sur l'éthique de l'IA et l'innovation responsable), des risques pour la paix et la sécurité découlant du détournement et de l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables;
- (b) établir des liens avec les étudiants - exposer aux étudiants en sciences, technologies, ingénierie et mathématiques (STEM) choisis à travers le monde la manière dont les risques pour la paix et la sécurité que font peser le détournement et l'utilisation abusive du développement civil de l'IA par des acteurs irresponsables peuvent être recensés, prévenus ou atténués dans le cadre du processus de recherche et d'innovation ou par d'autres processus de gouvernance; et
- (c) établir des liens avec l'industrie de l'IA - coopérer avec des associations professionnelles et des organismes de normalisation tels que l'Institut de l'ingénierie électrique et électronique (IEEE) pour i) diffuser du matériel pédagogique sur mesure et des activités de mobilisation auprès des techniciens; ii) soutenir les utilisations positives de l'IA au service de la paix et de la sécurité; et iii) faciliter le dialogue et le partage d'informations entre les experts du monde universitaire, du secteur privé et des pouvoirs publics sur la manière d'atténuer le risque de détournement et d'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables.

Une telle approche permet aux projets d'atteindre la communauté de l'IA à tous les niveaux, y compris non seulement les praticiens actuels, mais aussi les générations futures. Elle permet également une coopération par-delà les cloisonnements universitaires, sectoriels et autres, et favorise la durabilité des efforts futurs en mettant en place des réseaux qui transcendent ces limites.

Les projets visent également à tirer parti du pouvoir de mobilisation et l'expérience du SIPRI et du Bureau des affaires de désarmement pour influencer la communauté de l'IA à l'échelle mondiale, et non les seules parties prenantes de l'UE. Le SIPRI et le Bureau des affaires de désarmement des Nations unies sont les mieux placés pour atteindre les acteurs de l'IA issus de toute l'Afrique, de la région Asie-Pacifique, de l'Europe et de l'Amérique du Nord et du Sud et faciliter leur coopération. Les deux entités ont également une expérience professionnelle dans d'autres domaines scientifiques et technologiques confrontés à des défis similaires en matière de double usage et de prolifération, y compris la biotechnologie. Les projets visent également à tirer parti des conditions qui existent au sein de l'Union européenne, telles que a) l'existence de processus multipartites avancés en matière d'IA responsable; b) le niveau élevé de mobilisation et d'expertise en matière de désarmement, de maîtrise des armements et de non-prolifération dans l'UE; c) la diversité des liens que les organisations universitaires, de recherche et du secteur privé de l'UE entretiennent avec d'autres régions, notamment dans les pays du Sud, qui constitueront également un objectif majeur de la mobilisation; et d) la diversité des nationalités des étudiants, des éducateurs et des ingénieurs des universités, des instituts de recherche et du secteur privé.

L'inclusion sera un élément essentiel pour la conduite des activités des projets. Afin de soutenir efficacement la communauté de l'IA, les projets reconnaissent que la communauté de l'IA se compose d'un large éventail d'acteurs, et plus particulièrement que:

- (a) l'égalité entre les hommes et les femmes est un facteur très important. Aussi l'égalité entre les hommes et les femmes sera-t-elle intégrée conformément aux stratégies de parité et de prise en compte systématique de cette question à l'échelle du système des Nations unies. La participation des femmes à toutes les activités du projet sera encouragée et requise; et

- (b) l'inclusion des personnes handicapées et l'aménagement raisonnable seront assurés tout au long de la durée des projets. Il s'agira notamment de lever les obstacles à la participation des personnes handicapées et de veiller à ce que des mesures soient prises pour coopérer avec les personnes handicapées et faciliter la représentation de leurs points de vue et de leurs expériences sur des questions de fond.

3. PROJETS

Les trois projets décrits ci-dessous sont censés se compléter et se renforcer mutuellement, forts d'éléments qui s'appliquent tout au long des 36 mois.

Projet n° 1 - Production, à l'intention de la communauté civile de l'IA, de matériel pédagogique et portant sur le renforcement des capacités

3.1.Objectif du projet

Le projet n° 1 s'attache principalement à fournir aux acteurs civils de l'IA des connaissances et des moyens leur permettant d'évaluer et d'atténuer les risques que le détournement et l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables peuvent faire peser sur la paix et la sécurité internationales. Il vise à produire du matériel pédagogique et portant sur le renforcement des capacités qui fourniront aux praticiens de l'IA de toutes les régions, de tous les niveaux et de tous les secteurs (y compris les éducateurs qui s'occupent d'IA, les concepteurs de programmes d'études, les étudiants en sciences, technologies, ingénierie et mathématiques, les ingénieurs en IA et les chercheurs du monde universitaire et du secteur privé) les informations et les outils nécessaires pour:

- (a) comprendre comment la recherche et l'innovation civiles dans le domaine de l'IA pourraient être détournées et utilisées abusivement d'une manière qui pourrait présenter des risques pour la paix et la sécurité internationales et comment les décisions en matière de développement et de diffusion de la recherche et de l'innovation peuvent accroître ou réduire le risque de détournement et d'utilisation abusive;
- (b) comprendre les efforts déjà déployés par la communauté du désarmement, de la maîtrise des armements et de la non-prolifération pour atténuer les risques de détournement et d'utilisation abusive de la recherche et de l'innovation civiles; et
- (c) pratiquer une innovation responsable d'une manière qui atténue le risque de détournement et d'utilisation abusive dans le développement et la diffusion de la recherche et de l'innovation.

3.1.1. Description du projet

Ce projet produira trois ensembles distincts de matériel pédagogique et portant sur le renforcement des capacités.

- (a) Manuel (1) - Le manuel fera l'inventaire des connaissances de base et des moyens permettant aux acteurs de l'IA d'évaluer et d'atténuer, dans le processus de recherche et d'innovation, les risques de détournement et d'utilisation abusive des technologies civiles dans le domaine de l'IA par des acteurs irresponsables. Il examinera pourquoi et comment les décisions relatives au développement et à la diffusion de la recherche et de l'innovation peuvent avoir une incidence sur les risques de détournement et d'utilisation abusive et, à leur tour, faire naître des risques ou des possibilités pour la paix et la sécurité. Il instituera en outre des obligations pertinentes en matière de droit international et de contrôle des exportations, en sus des considérations de sûreté et de sécurité examinées dans les milieux militaires ainsi que dans les milieux du désarmement, de la maîtrise des armements et de la non-prolifération, et présentera des exemples de processus et d'outils permettant de mettre en œuvre une innovation responsable, par exemple des méthodes d'analyse d'impact technologique et des modèles d'évaluation des risques.

- (b) *Série de podcasts (~ 10)* - Ces podcasts serviront de support accessible et attrayant pour permettre aux acteurs de l'IA d'apprendre pourquoi et comment des processus d'innovation responsables dans le domaine de l'IA peuvent soutenir la paix et la sécurité internationales en atténuant les risques que présentent le détournement et l'utilisation abusive par acteurs irresponsables. Cette série abordera des thèmes importants (par exemple, le modèle de détournement et d'utilisation abusive de la recherche et de l'innovation à double usage et à usage général; les défis humanitaires, stratégiques et politiques liés à l'utilisation potentiellement abusive de la recherche et de l'innovation civiles dans le domaine de l'IA; les défis auxquels sont confrontés les milieux du désarmement, de la maîtrise des armements et de la non-prolifération dans le cadre des efforts d'atténuation des risques; comment mener une innovation responsable grâce à l'évaluation des risques; le respect des règles en matière de contrôle des exportations; la réduction des risques dès le stade de la conception; l'édition responsable; connaître vos clients; et l'expérience issue des exercices de simulation) et sera structurée autour d'entretiens que l'équipe du projet mènera avec des représentants des communautés concernées.
- (c) *Série de billets de blog (9-10)* - L'équipe élaborera une série de billets de blog à contenu édité visant à accroître la visibilité des efforts destinés à transcender les limites qui séparent la communauté de l'IA responsable axée sur les efforts d'experts issus de la sphère civile et la communauté de la maîtrise des armements et de la non-prolifération. La série de billets de blog constituera une plateforme permettant de diffuser des informations, des idées et des solutions concernant l'identification et la gestion des risques liés au détournement et à l'utilisation abusive de l'IA civile dans le processus de recherche et d'innovation. Le blog s'efforcera de représenter la diversité des idées et des perspectives présentes dans le secteur de l'IA.

Ces supports seront diffusés publiquement grâce aux sites web des acteurs de mise en œuvre, à leur présence sur les médias sociaux et à une communication directe avec les entités universitaires concernées, les associations professionnelles civiles dans le domaine de l'IA et d'autres groupes appropriés.

3.1.2. Résultats attendus du projet

Ce projet devrait permettre d'établir un nouvel ensemble de supports permettant aux praticiens civils de l'IA de prendre conscience a) de la façon dont la recherche et l'innovation civiles dans le domaine de l'IA pourraient être détournées et utilisées abusivement d'une manière qui pourrait présenter des risques pour la paix et la sécurité internationales, b) de la façon dont ces risques sont traités par la communauté du désarmement, de la maîtrise des armements et de la non-prolifération, et c) de la façon dont les praticiens de l'IA pourraient contribuer davantage à atténuer ces risques au moyen de processus d'innovation responsables.

Cela devrait permettre de renforcer la mobilisation du secteur civil de l'IA en matière d'atténuation des risques que le détournement et l'utilisation abusive de l'IA civile peuvent faire peser sur la paix et la sécurité internationales, et d'améliorer la capacité des praticiens techniques à s'investir dans les processus pertinents de la communauté du désarmement, de la maîtrise des armements et de la non-prolifération, ainsi que de contribuer à la mobilisation de nouveaux publics qui ne sont généralement pas inclus dans les efforts d'éducation au désarmement et à la non-prolifération.

Ce projet devrait également permettre de soutenir la mise en œuvre des autres projets et servira de base aux activités pédagogiques et portant sur le renforcement des capacités menées dans le cadre du projet n° 2, ainsi qu'aux activités de dialogue et de mobilisation menées dans le cadre du projet n° 3. Ces activités devraient à leur tour contribuer à la production et à l'amélioration des supports. Une telle approche itérative devrait contribuer à lever les obstacles potentiels à leur promotion, à leur diffusion et à leur utilisation au sein de la communauté de l'IA, y compris pour ce qui a trait aux questions liées à la langue, au contenu, au contexte et à la disponibilité,

obstacles qui pourraient nuire à leur efficacité au niveau mondial, en particulier dans les pays du Sud.

3.2. Projet n° 2 - Mise en place, à l'intention des futurs praticiens de l'IA, d'activités pédagogiques et portant sur le renforcement des capacités

3.2.1. Objectif du projet

L'objectif du projet n° 2 est de favoriser l'intégration, dans la formation des générations à venir de praticiens de l'IA, du problème du détournement et de l'utilisation abusive de la recherche civile sur l'IA par des acteurs irresponsables. À terme, cela permettra de faire en sorte que les étudiants en sciences, technologies, ingénierie et mathématiques (STEM), qui façonnent l'avenir de l'IA, soient conscients des incidences négatives que le détournement et l'utilisation abusive de leur travail par des acteurs irresponsables pourraient avoir sur la paix et la sécurité internationales et qu'ils disposent des outils de base nécessaires pour identifier et atténuer un tel risque dans le processus de recherche et d'innovation.

Dans le contexte de ce projet, une série d'ateliers pédagogiques et portant sur le renforcement des capacités seront organisés avec des éducateurs et des étudiants, en collaboration avec des universités et des acteurs industriels internationaux sélectionnés. Le projet vise ainsi à mettre en place des activités de renforcement des capacités auxquelles les éducateurs et les concepteurs de programmes universitaires pourraient recourir pour inclure dans la formation des futurs praticiens de l'IA (via des cours sur l'éthique et l'innovation responsable dans le domaine de l'IA, par exemple) des considérations relatives aux risques de détournement et d'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA par des acteurs irresponsables et de relier ces risques au contexte plus large de paix et de sécurité. Dans le cadre de ces ateliers, le projet visera également à recenser un réseau d'éducateurs, de concepteurs de programmes d'études et d'étudiants intéressés qui soutiendraient la diffusion et la promotion, au sein de la communauté de l'éducation à l'IA et de la communauté des praticiens de l'IA, du matériel pédagogique et des activités portant sur le renforcement des capacités prévus par le projet. Ce volet de mise en réseau vise à assurer la viabilité des projets au delà de leur durée immédiate et à permettre l'établissement de liens plus étroits en vue de soutenir une mobilisation des milieux techniques civils en faveur d'objectifs plus larges en matière de paix, de sécurité, de désarmement et de maîtrise des armements.

3.2.2. Description du projet

Dans le contexte de ce projet, une série d'ateliers pédagogiques et portant sur le renforcement des capacités seront organisés avec des éducateurs et des étudiants issus d'une sélection d'universités à travers le monde. Il s'agirait d'une combinaison de conférences et d'activités interactives qui offriront aux éducateurs et aux étudiants la possibilité de réfléchir à la façon dont la recherche et l'innovation civiles en matière d'IA pourraient être détournées et utilisées abusivement d'une manière qui pourrait présenter des risques pour la paix et la sécurité internationales et à la façon dont ces risques peuvent être identifiés, prévenus ou atténués dans le cadre du processus de recherche et d'innovation, ou par d'autres processus de gouvernance. Ces activités s'appuieront sur des travaux pilotes menés précédemment à plus petite échelle par le Bureau des affaires de désarmement des Nations unies, qui a expérimenté des méthodes visant à prendre contact avec des étudiants en sciences, technologies, ingénierie et mathématiques et à les sensibiliser à l'importance qu'il y a à tenir compte de l'impact plus large de leur travail et à s'intéresser à l'expertise en dehors de leur domaine d'origine. Concrètement, les activités en question comprendraient:

- (a) *des ateliers régionaux de renforcement des capacités à l'intention des éducateurs et des étudiants (4)* - ces ateliers régionaux mèneront et promouvront des activités auxquelles les éducateurs peuvent recourir pour renforcer les capacités des étudiants en sciences, technologies, ingénierie et mathématiques en matière d'innovation responsable dans le

domaine de l'IA, l'accent étant mis en particulier sur la manière d'évaluer et d'atténuer les risques de détournement et d'utilisation abusive des technologies civiles dans le domaine de l'IA par des acteurs irresponsables. Chaque atelier sera organisé avec une université établie dans l'UE et une grande université d'une région différente du monde, ce qui devrait permettre de toujours mettre en relation un ensemble diversifié de participants situés dans l'UE et un ensemble diversifié de participants situés en dehors de l'UE. Les ateliers couvriront ensuite l'Amérique latine et les Caraïbes, l'Amérique du Nord, l'Afrique et l'Asie et le Pacifique. Cela permettra la participation d'étudiants (aux niveaux maîtrise et doctorat) du monde entier, y compris du Sud. Les ateliers se dérouleront principalement en anglais, mais, dans la mesure du possible, les participants auront la possibilité de participer à des activités axées sur d'autres groupes linguistiques; et

- (b) un atelier international sur le renforcement durable des capacités (1) - cet atelier s'appuiera sur les enseignements tirés des ateliers régionaux et facilitera l'échange d'informations et d'expériences entre les éducateurs et certains étudiants des universités participant au projet. L'atelier s'attachera à examiner la manière d'affiner les activités et les outils élaborés tout au long du projet et de les diffuser au delà du groupe d'universités participantes. Il s'attachera également à examiner la manière d'encourager les étudiants à s'investir, une fois qu'ils seront entrés sur le marché du travail, dans une IA responsable qui traite les risques que font peser le détournement et l'utilisation abusive sur la paix et la sécurité internationales.

Les réseaux et la présence du SIPRI et du Bureau des affaires de désarmement des Nations unies en Afrique, dans la région Asie-Pacifique, en Europe, en Amérique du Nord et en Amérique du Sud seront mis à profit pour faciliter et soutenir certains aspects des activités, en tant que de besoin.

3.2.3. Résultats attendus du projet

Le projet devrait créer des modèles d'activités de renforcement des capacités et de mobilisation que les éducateurs et les concepteurs de programmes universitaires pourraient reproduire pour sensibiliser les futurs praticiens de l'IA aux problèmes du détournement et de l'utilisation abusive de l'IA civile par des acteurs irresponsables et à la manière dont ils peuvent contribuer à atténuer ces problèmes au moyen de processus d'innovation responsables. Une fois les activités du projet achevées, les participants (éducateurs, mais aussi étudiants en sciences, technologies, ingénierie et mathématiques) devraient être en mesure d'utiliser et de promouvoir des outils, méthodes et concepts responsables en matière d'innovation pour identifier et atténuer les risques de détournement et d'utilisation abusive dans le développement et la diffusion de la recherche et de l'innovation civiles dans le domaine de l'IA.

Les activités du projet devraient également permettre de créer un réseau d'éducateurs, de concepteurs de programmes d'études et d'étudiants qui non seulement promouvraient les activités du projet au sein des communautés de l'éducation et de professionnels dans le domaine de l'IA (par exemple lors de conférences de la Computational Intelligence Society de l'IEEE), mais seraient également disponibles pour contribuer, capacités techniques à l'appui, aux processus de gouvernance internationale menés sous l'égide des États (par exemple, le processus de la convention sur l'emploi de certaines armes conventionnelles sur les technologies émergentes dans le domaine des systèmes d'armes létales autonomes).

L'utilité à court et à long terme de ces activités sera démontrée au moyen d'enquêtes ex ante et ex post.

3.3. Projet n° 3 - Faciliter le développement durable, la diffusion et l'impact à long terme de l'innovation responsable dans le domaine de l'IA au service de la paix et de la sécurité

3.3.1. Objectif du projet

Le projet n° 3 a pour objectif de faciliter le développement durable, la diffusion et l'impact à long terme de l'innovation responsable dans le domaine de l'IA en tant que moyen d'atténuer les risques que le détournement et l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA peuvent représenter pour la paix et la sécurité. Il entend y parvenir par des tables rondes avec le secteur de l'IA, des dialogues multipartites, l'établissement d'un rapport public et des activités de diffusion ciblées. Le projet vise à faire en sorte que le travail généré, en particulier les activités pédagogiques, de renforcement des capacités et de mobilisation, touche et influence la communauté de l'IA dans son ensemble, à tous les niveaux (des étudiants aux ingénieurs et autres professionnels de l'IA) et par delà les limites géographiques, sectorielles et autres. Pour accroître les possibilités de produire un impact large et profond, il est essentiel de coopérer avec les organisations professionnelles dans ce domaine, telles que l'IEEE, et de mener une coopération multidimensionnelle allant au delà les cloisonnements universitaires, sectoriels et autres. Ces efforts permettront aux représentants intéressés de différentes communautés de l'IA de prendre en charge le problème et de donner leurs propres points de vue sur la manière dont les efforts d'atténuation des risques peuvent être menés et promus de manière durable au sein et dans l'ensemble de la communauté mondiale de l'IA. Il est également important, du point de vue de l'utilité à long terme du projet pour les États, les organisations intergouvernementales et d'autres acteurs, que les praticiens de l'IA puissent apprendre au contact des experts gouvernementaux participant à l'atténuation des risques dans le contexte du désarmement, de la maîtrise des armements et de la non-prolifération et coopérer avec eux. Il est également essentiel, pour la durabilité, de veiller à ce que les connaissances générées par les activités de mobilisation soient analysées, consolidées et diffusées de manière appropriée.

3.3.2. Description du projet

Ce projet comprend les principaux volets suivants:

- (a) des dialogues multipartites sur "Une innovation responsable en matière d'IA au service de la paix et de la sécurité" (9 au maximum) - cette série de réunions virtuelles de dialogue rassemblerait des experts du monde universitaire, des milieux de la recherche, du secteur privé et du secteur de la maîtrise des armements classiques de l'UE et d'ailleurs afin d'examiner les points suivants:
 - i. les tendances technologiques susceptibles de faire naître des risques en matière de diffusion, de détournement et d'utilisation abusive, avec des répercussions sur la paix et la sécurité internationales;
 - ii. la façon de s'investir dans l'atténuation des risques au moyen de processus, de méthodes et de moyens d'innovation responsables, ainsi que les possibilités et les défis en matière de dialogue et de partage des connaissances entre les communautés de parties prenantes, y compris celles qui opèrent dans d'autres secteurs tels que ceux de la biologie et de la chimie; et
 - iii. l'utilité, la finalité et le format potentiels d'un réseau autonome d'experts et d'activités de dialogue. Le groupe d'experts se réunira plusieurs fois par an et travaillera à l'organisation de deux événements publics pour la communauté au sens large.

Sur les neuf réunions virtuelles, deux sont destinées à être ouvertes au public afin de faciliter une consultation plus large.

- (b) des tables rondes avec le secteur privé (6 au maximum) - cette série de tables rondes virtuelles permettra d'entamer avec les acteurs tirant parti de processus d'innovation responsables en matière d'IA dans le secteur privé (par exemple, le partenariat sur l'IA) un dialogue sur la manière dont ils peuvent contribuer à réduire au minimum les risques de détournement et d'utilisation abusive des technologies civiles dans le domaine de l'IA par des acteurs irresponsables, et étudiera d'éventuelles incitations dans le cadre du développement du secteur privé en vue de la réalisation de cet objectif. Les thèmes abordés seront notamment les suivants:
- i. l'importance du contexte international en matière de sécurité et de désarmement pour le secteur privé;
 - ii. le ou les environnements juridiques dans lesquels le développement, le déploiement et l'exploitation de l'IA s'inscrivent à travers le monde;
 - iii. la manière de s'appuyer sur des mécanismes d'évaluation des risques et d'autres mesures qui font parti de processus d'innovation responsables et de programmes en matière de conformité des entreprises, ou pourraient y être intégrés; et
 - iv. les enseignements à tirer d'autres secteurs, processus et cadres liés à la maîtrise des armements (par exemple, les secteurs de la biologie et de la chimie).
- (c) un rapport sur les points de vue de la communauté de l'IA en matière de maîtrise des armements et d'atténuation des risques dans le domaine de l'IA (I), axé sur les communautés de l'IA et de la maîtrise des armements (I) - l'élaboration de ce rapport intégrera et consolidera les conclusions et les recommandations du projet dans un document de référence unique destiné à la fois à la communauté de l'IA responsable axée sur les efforts d'experts issus de la sphère civile et la communauté de la maîtrise des armements. Le rapport examinerait la manière dont les risques internationaux pour la paix et la sécurité liés au détournement et à l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA peuvent être identifiés, évalués et traités.
- (d) des événements de diffusion axés sur la consultation et le dialogue avec les communautés de l'IA et de la maîtrise des armements (à déterminer) - l'équipe cherchera des occasions de communiquer les travaux, leurs conclusions et leurs recommandations tout au long du projet. Le format des événements et le contenu des présentations seront adaptés aux besoins des groupes cibles. Il peut s'agir de réunions du CONOP, de l'Assemblée de l'Alliance européenne pour l'IA ; du groupe d'experts gouvernementaux sur les technologies émergentes dans le domaine des systèmes d'armes létales autonomes; du groupe de travail interinstitutions sur l'IA (IAWG-AI); de l'initiative de l'UIT "L'IA au service du bien social"; du dialogue annuel sur l'innovation mené dans le cadre de l'Institut des Nations unies pour la recherche sur le désarmement (UNIDIR); et de l'Institut de l'ingénierie électrique et électronique. L'équipe s'efforcera également de nouer des contacts bilatéraux avec les parties prenantes concernées des pouvoirs publics, du monde universitaire et du secteur privé.

3.3.3. Résultats attendus du projet

Ce projet devrait jeter les bases du développement durable, de la diffusion et de l'impact de l'innovation responsable des processus d'IA visant à lutter contre les risques de diffusion, de détournement et d'utilisation abusive et leurs implications pour la paix et la sécurité au delà de la durée immédiate de la décision du Conseil.

Le dialogue multipartite devrait servir de modèle pour le partage d'informations et la collaboration en matière d'atténuation des risques, non seulement au sein de la communauté mondiale de l'IA, mais aussi entre la communauté de l'IA responsable axée sur les efforts d'experts issus de la sphère civile et les communautés du désarmement, de la maîtrise des armements et de la non-prolifération. Un tel modèle pourrait être utilisé pour familiariser les décideurs politiques avec les principales avancées technologiques et scientifiques concernant

l'innovation responsable dans le domaine de l'IA et pour familiariser les publics techniques avec l'environnement dans lequel les décideurs s'investissent à l'heure actuelle. Le projet devrait faciliter des relations et une mobilisation durables entre les acteurs intéressés au sein de ces différentes communautés et entre celles-ci. Ces effets de réseau hétérogènes devraient permettre un développement plus important et une promotion généralisée de l'innovation responsable dans le domaine de l'IA au service de la paix et de la sécurité au delà de la durée du projet.

Le dialogue avec le secteur privé devrait permettre au secteur privé de l'IA de participer davantage et de manière plus étroite à l'identification, à la prévention et à l'atténuation des risques pour la paix et la sécurité découlant du détournement et de l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA. Le projet devrait permettre aux principaux acteurs des processus ayant trait au secteur privé de mieux comprendre et de mieux prendre en charge les problèmes auxquels il cherche à faire face. En outre, il vise à faciliter l'adoption et la mise en œuvre (plus larges) de processus, de méthodes et de moyens d'innovation responsables dans les mécanismes et procédures existants de gestion des risques d'entreprise.

Le dialogue multipartite et la table ronde avec le secteur privé devraient également permettre de recueillir des informations sur un certain nombre de questions importantes, notamment a) la façon dont les méthodes et les moyens d'innovation responsables peuvent être affinés et déployés pour identifier, prévenir et atténuer les risques liés au détournement et à l'utilisation abusive de la recherche et de l'innovation civiles dans le domaine de l'IA; b) la façon dont la recherche et l'innovation en matière d'IA peuvent être utilisées de manière positive pour soutenir les objectifs de paix et de sécurité (grâce par exemple à des applications relatives à l'alerte précoce en cas de conflit et à l'aide humanitaire); et c) la façon de faciliter le dialogue et le partage d'informations entre les efforts d'atténuation des risques entrepris par la communauté de l'IA responsable axée sur l'usage civil (c'est-à-dire diverses initiatives menées sous l'égide de l'IEEE) avec ceux déjà en cours dans la communauté du désarmement, de la maîtrise des armements et de la non-prolifération au niveau intergouvernemental.

Le rapport et les activités de diffusion analyseront, consolideront et diffuseront les informations obtenues dans le cadre des projets n° 1, 2 et 3, et soutiendront ainsi la mise en avant des conclusions et recommandations des activités du projet au sein de la communauté mondiale de l'IA, ainsi qu'au sein des pouvoirs publics. Ils devraient également contribuer à assurer la durabilité de l'impact au delà de la durée des projets.

4. DURÉE

La durée totale de la mise en œuvre des projets est estimée à trente-six mois.