



Съвет на
Европейския съюз

Брюксел, 26 октомври 2022 г.
(OR. en)

14095/22
ADD 1

LIMITE

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

ПРЕДЛОЖЕНИЕ

От:	Върховния представител на Съюза по въпросите на външните работи и политиката на сигурност, подписано от г-н Stefano SANNINO, генерален секретар
Дата на получаване:	26 октомври 2022 г.
До:	Г-жа Thérèse BLANCHET, генерален секретар на Съвета на Европейския съюз
Относно:	Предложение на върховния представител на Съюза по въпросите на външните работи и политиката на сигурност до Съвета за решение на Съвета в подкрепа на изпълнението на проект „Насърчаване на отговорни иновации в областта на изкуствения интелект за мир и сигурност“

Приложено се изпраща на делегациите документ HR(2022) 238.

Приложение: HR(2022) 238

ЕВРОПЕЙСКА СЛУЖБА ЗА ВЪНШНА ДЕЙНОСТ



ПРОЕКТЕН ДОКУМЕНТ

Насърчаване на отговорни иновации в областта на ИИ за мир и сигурност

HR(2022)238

1. КОНТЕКСТ

Неотдавнашният напредък в областта на изкуствения интелект (ИИ) откри нови възможности за подкрепа и поддържане на мира и сигурността, например чрез технологични подобрения в области като ранното предупреждение за конфликти и проверката на контрола върху оръжията и износа. От друга страна, този напредък даде възможност за нови начини за създаване или задълбочаване на вече съществуващо напрежение, конфликти и несигурност между и в рамките на държавите. Рисковете, свързани с някои приложения на ИИ, като например смъртоносните автономни оръжейни системи, се откриха като основен проблем, предизвикващ загриженост за общността за контрол върху оръжията. Една рискова възможност, на която трябва да се обърне по-голямо внимание и за която настоящите усилия за контрол върху оръжията и дипломатическите усилия може да не са достатъчни, за да се реагира, е отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях от безотговорни участници, включително злонамерени недържавни участници, като например злоупотребата с генеративно-състезателни мрежи за създаване на дълбоки фалшификати („дийпфейк“) за кампании за дезинформация.

ИИ е базова технология с голям потенциал за обща употреба. До научните изследвания и иновациите в областта на ИИ, разработвани за граждански приложения, може (относително лесно) да се получи достъп и те да бъдат пренасочени от някои участници за вредни или разрушителни употреби, които да имат последици за международния мир и сигурност. Отклоняването на граждански технологии и злоупотребата с тях не са нови явления, нито са ограничени само до ИИ. В близката област на роботиката това наскоро беше илюстрирано от използването на безпилотни летателни апарати за развлекателни цели като оръжие от Даиш/ИДИЛ в Сирия. Но в случая с ИИ проблемът се усложнява от множество фактори: нематериалното и бързо променящо се естество на алгоритмите и данните от областта на ИИ, което затруднява контрола върху трансфера/разпространението на тези алгоритми и данни; водещата роля на частния

сектор в екосистемата за научни изследвания, развойна дейност и иновации и произтичащата от това необходимост от защита на патентованите алгоритми; както и наличието в световен мащаб на човешки експертен опит и материални ресурси, способни да променят предназначението на ИИ технологиите. Междувременно лицата, работещи в областта на ИИ в гражданския сектор, твърде често не са наясно с потенциалните последици, които отклоняването на тяхната работа и злоупотребата с нея биха могли да имат за международния мир и сигурност, или не проявяват желание да участват в съществуващите дискусии относно рисковете от ИИ в кръговете на контрола върху оръжията и неразпространението на оръжия.

Необходимо е да се подкрепи по-активното участие на гражданската ИИ общност в разбирането и смекчаването на рисковете за мира и сигурността, свързани с отклоняването на граждански ИИ технологии и злоупотребата с тях от безотговорни участници. За Стокхолмския международен институт за изследване на проблемите на мира (SIPRI) и Службата на ООН по въпросите на разоръжаването (UNODA) това би могло да се постигне чрез по-голямо насърчаване на отговорни иновации в световната гражданска ИИ общност. Досегашната работа на SIPRI и UNODA показва, че отговорните иновации, като механизъм за самоуправление, биха могли да предоставят на световната гражданска ИИ общност практически инструменти и методи за идентифициране и за подпомагане на предотвратяването и смекчаването на рисковете, които отклоняването и злоупотребата с граждански научни изследвания и иновации в областта на ИИ биха могли да създадат за мира и сигурността. Работата на SIPRI и UNODA позволи също така да се установят методики и няколко текущи граждански инициативи за отговорен ИИ, които биха могли да се доразвият за повишаване на чувствителността на гражданската ИИ общност към въпросите, свързани с контрола върху оръжията и неразпространението на оръжия, експертните дебати и държавните позиции относно отговорното развитие, разпространението и използването на ИИ, както и поуките, които трябва да бъдат извлечени от работата, свързана с отговорността в сектора на отбраната.¹ От решаващо значение е, че тази предварителна работа ясно определи взаимодействието със студентите в областта на науките, технологиите, инженерството и математиката (НТИМ), които работят с ИИ все още в образователен формат, като централен елемент за всяко ефективно усилие за отговорни иновации.

2. ЦЕЛИ

Тези проекти са насочени към подкрепа на по-активното участие на гражданската ИИ общност за смекчаване на рисковете, които отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях от безотговорни

¹ Методиките включват например препоръчителните практики на Института на инженерите по електротехника и електроника (IEEE) за оценка на въздействието на автономните и интелигентните системи върху благосъстоянието на хората (IEEE Std 7010-2020) и списъка на експертната група на високо равнище по въпросите на ИИ за оценка на надеждността на изкуствения интелект (ALTAI). Инициативите включват: Глобалната инициатива на IEEE за етика на автономните и интелигентните системи; партньорството в областта на ИИ; глобалното партньорство в областта на ИИ.

участници могат да представляват за международния мир и сигурност. Те имат за цел да постигнат това, първо, като предизвикат по-добро разбиране за това как решенията за разработване и разпространение на научни изследвания и иновации в областта на ИИ могат да окажат въздействие върху рисковете от отклоняване и злоупотреба и на свой ред да създадат риск или възможности за мир и сигурност, и второ, чрез насърчаване на процеси, методи и инструменти за отговорни иновации, които да могат да спомогнат за гарантиране на мирното прилагане на гражданските иновации и отговорното разпространение на знания в областта на ИИ. За тази цел те подкрепят дейности за изграждане на капацитет, научни изследвания и взаимодействие, които *i)* ще увеличат капацитета в рамките на световната гражданска ИИ общност за включване и преодоляване на рисковете за мира и сигурността, свързани с отклоняването на граждански ИИ и злоупотребата с него от безотговорни участници чрез процеси за отговорни иновации; и *ii)* ще укрепят връзката между усилията за смекчаване на риска в областта на отговорния ИИ в гражданската сфера и усилията, които вече са в ход в общността за разоръжаване, контрол на оръжията и неразпространение на оръжия на междуправителствено равнище. От решаващо значение е, че те не възнамеряват да установяват нови стандарти, принципи или разпоредби, нито да се намесват по друг начин в области от компетентността на държавите. Вместо това те възнамеряват да развият граждански усилия за отговорни иновации, за да включат рисковете за мира и сигурността, породени от отклоняването на граждански ИИ и злоупотребата с него от безотговорни участници, и да осигурят образование относно съществуващите съответни междуправителствени усилия.

За да достигнат ефективно до гражданската ИИ общност и да ѝ окажат въздействие, проектите прилагат подход, основан на три направления, който има за цел:

- (a) взаимодействие с преподавателите — работа с избрани преподаватели и съставители на академични учебни програми относно разработването и популяризирането на образователни материали, които могат да се използват за интегриране в обучението на бъдещи специалисти в областта на ИИ (например в рамките на курсове по етика на ИИ и отговорни иновации) на въпроса за рисковете за мира и сигурността, произтичащи от отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях от страна на безотговорни участници;
- (b) взаимодействие със студентите — запознаване на избрани студенти в областта на науките, технологиите, инженерството и математиката (НТИМ) от цял свят с начините за идентифициране на рисковете за мира и сигурността, породени от отклоняването на граждански разработки на ИИ и злоупотребата с тях от страна на безотговорни участници, тяхното предотвратяване или намаляване в процеса на научните изследвания и иновациите или чрез други процеси на управление; и
- (c) взаимодействие със сектора на ИИ — работа с професионални сдружения и органи по стандартизация като Института на електроинженерите и инженерите по електроника (IEEE) за *i)* разпространение на адаптирани образователни материали и дейности за сътрудничество сред техническите специалисти; *ii)* подкрепа за положителното използване на ИИ за целите на мира и сигурността; и *iii)*

улесняване на диалога и обмена на информация между експерти от академичните среди, частния и държавния сектор относно начините за намаляване на риска от отклоняване на граждански научни изследвания и иновации в областта на ИИ и злоупотреба с тях от страна на безотговорни участници.

Този подход позволява на проектите да достигнат до ИИ общността на всички равнища, включително не само настоящите специалисти, но и бъдещите поколения. Той също така дава възможност за взаимодействие между академичните среди, промишлеността и други сектори и подкрепя устойчивостта на бъдещите усилия чрез създаване на мрежи, които излизат извън тези граници.

Целта на проектите е да се използват и обединяващата сила и опит на SIPRI и UNODA, за да се окаже въздействие върху ИИ общността в световен мащаб, а не само върху заинтересованите страни от ЕС. SIPRI и UNODA са в изключителната позиция да могат да достигнат до участниците в областта на ИИ от цяла Африка, Азиатско-тихоокеанския регион, Европа и Северна и Южна Америка и да улеснят взаимодействието между тях. И двете организации имат опит и в други области на науките и технологиите, изправени пред сходни предизвикателства на двойна употреба и разпространение, включително биотехнологиите. Стремешът е проектите да се възползват и от съществуващите условия в рамките на Европейския съюз, като а) наличието на напреднали процеси относно отговорния ИИ с участието на множество заинтересовани страни; б) високото равнище на ангажираност и експертен опит по въпросите на разоръжаването, контрола върху оръжията и неразпространението в ЕС; в) разнообразието от връзки на академичните и научноизследователските организации и организациите от частния сектор в ЕС с други региони, по-специално в Глобалния Юг, което също ще бъде основна цел за взаимодействие; и г) многообразието от националности на студентите, преподавателите и инженерите в университетите, научноизследователските институти и частния сектор.

Приобщаването ще бъде основно съображение за провеждането на дейностите по проектите. За да се подкрепи ефективно ИИ общността, в проектите се отчита, че тази общност се състои от широк спектър участници, и по-специално, че

- (a) полът е изключително важен фактор. Поради тази причина равенството между половете ще бъде интегрирано в съответствие със стратегиите в цялостната система на ООН за интегриране на принципа на равенство между половете и равнопоставеност. Участието на жените във всички дейности по проекта ще бъде насърчавано и изисквано; и
- (b) през цялото време ще се практикува приобщаването на хората с увреждания и разумното приспособяване към нуждите. Това ще включва преодоляване на пречките пред участието на хората с увреждания, както и гарантиране, че се предприемат стъпки за взаимодействие с хората с увреждания и за улесняване на представянето по същество на техните гледни точки и опит.

3. ПРОЕКТИ

Трите проекта, описани по-долу, следва да се допълват и подкрепят взаимно, като елементите продължават общо 36 месеца.

Проект 1 — Изготвяне на образователни материали и материали за изграждането на капацитет за гражданската ИИ общност

3.1.Цел на проекта

Проект 1 е насочен към предоставяне на знания и средства, чрез които гражданските участници в областта на ИИ да оценяват и намаляват рисковете, които отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях от безотговорни участници могат да представляват за международния мир и сигурност. Той има за цел изготвянето на образователни материали и материали за изграждането на капацитет, които ще предоставят на специалистите в областта на ИИ от всички региони, равнища и сектори (включително специализираните в областта на ИИ преподаватели, съставителите на учебни програми, студентите в областта на НТИМ и инженерите и изследователите в областта на ИИ от академичните среди и частния сектор) информацията и инструментите, необходими за

- (a) разбиране на механизмите, чрез които гражданските научни изследвания и иновации в областта на ИИ могат да бъдат отклонявани и използвани по начини, които биха могли да представляват риск за международния мир и сигурност, и как решенията за разработване и разпространение на научни изследвания и иновации могат да увеличат или намалят риска от отклоняване и злоупотреба;
- (b) разбиране на усилията, които е положила общността за разоръжаване, контрол върху оръжията и неразпространение на оръжия за намаляване на рисковете от отклоняване на граждански научни изследвания и иновации и злоупотреба с тях; и
- (c) реализиране на отговорни иновации по начин, който намалява риска от отклоняване и злоупотреба при разработването и разпространението на научни изследвания и иновации.

3.1.1. Описание на проекта

В рамките на този проект ще бъдат изготвени три отделни набора от образователни материали и материали за изграждането на капацитет.

- (a) *Наръчник (1)* — наръчникът ще съдържа основни знания и инструменти, с помощта на които в процеса на научните изследвания и иновациите участниците в областта на ИИ ще могат да оценяват и намаляват рисковете от отклоняване на граждански технологии с ИИ и злоупотреба с тях от безотговорни участници. В него ще се разглежда и въпросът защо и как решенията относно разработването и разпространението на научните изследвания и иновациите могат да оказват въздействие върху рисковете от отклоняване и злоупотреба и на свой ред да създават рискове или възможности за мира и сигурността. Той ще представя и

съответните задължения съгласно международното право и по отношение на контрола на износа в допълнение към съображенията за безопасност и сигурност, които се обсъждат в кръговете на военните и сред специалистите в областта на разоръжаването, контрола върху оръжията и неразпространението на оръжия; в него ще бъдат посочени и примери на процеси и инструменти за реализиране на отговорни иновации, като методики за оценка на въздействието на технологиите и образци за оценка на риска.

- (b) Подкастове (~10) — тези подкастове ще действат като достъпна и ангажираща среда, за да могат участниците в областта на ИИ да научат защо и как процесите за отговорни иновации в областта на ИИ могат да подкрепят международния мир и сигурност чрез смекчаване на рисковете, свързани с отклоняването и злоупотребата от безотговорни участници. Поредицата ще разгледа важни теми (например модела на отклоняване и злоупотреба с научни изследвания и иновации с двойна/обща употреба; хуманитарните, стратегическите и политическите предизвикателства, свързани с потенциалната злоупотреба с граждански научноизследователски иновации в областта на ИИ; предизвикателствата, с които се сблъскват кръговете по разоръжаване, контрол върху оръжията и неразпространението на оръжия в усилията за смекчаване на риска; начините за извършване на отговорни иновации чрез оценка на риска; съответствието с контрола на износа; намаляването на риска още при проектирането; отговорното публикуване; познаването на собствените клиенти; и опита от симулационни учения) и ще бъде структурирана около събеседвания, които екипът по проекта ще проведе с представители на съответните общности.
- (c) Поредица блогове (9–10) — екипът ще разработи подбрана поредица от публикации в блогове, чиято цел е да се повиши видимостта на усилията, които целят да пресекат границите между общността на насочения към граждански приложения „отговорен ИИ“ и общностите за контрол върху оръжията и неразпространение на оръжия. Поредицата блогове ще предостави платформа за разпространение на познания, идеи и решения относно идентифицирането и преодоляването на рисковете, свързани с отклоняването на граждански ИИ в процеса на научни изследвания и иновации и злоупотребата с него. Блогът ще има за цел да представи разнообразието от мисли и гледни точки в сектора на ИИ.

Тези материали ще се разпространяват публично чрез уебсайтовете на изпълнителите на проекта, присъствието им в социалните медии и чрез пряка комуникация със съответните академични субекти, граждански професионални сдружения в областта на ИИ и други подходящи групи.

3.1.2. Очаквани резултати от проекта

Очаква се този проект да създаде нов набор от материали, чрез които гражданските специалисти в областта на ИИ да бъдат информирани за: а) начините, по които гражданските научни изследвания и иновации в областта на ИИ биха могли да бъдат отклонени и използвани неправомерно така, че да представляват риск за международния

мир и сигурност, б) как тези рискове се разглеждат от общността по разоръжаване, контрол върху оръжията и неразпространение на оръжия и в) как специалистите в областта на ИИ биха могли допълнително да допринесат за смекчаване на тези рискове чрез процеси за отговорни иновации.

Очаква се това да ускори ангажираността на гражданския сектор в областта на ИИ за смекчаване на рисковете, които отклоняването на граждански ИИ и злоупотребата с него може да породи за международния мир и сигурност; подобряване на капацитета на техническите специалисти да се ангажират със съответните процеси в общността за разоръжаване, контрол върху оръжията и неразпространение на оръжия, както и подкрепа за привличането на нова публика, която традиционно не е включена в усилията за образование в областта на разоръжаването и неразпространението.

Очаква се също така материалите да подпомогнат изпълнението на другите проекти и да послужат като основа за образователните дейности и дейностите за изграждане на капацитет по проект 2, както и за дейностите за диалог и взаимодействие по проект 3. Очаква се тези дейности на свой ред да допринесат за изготвянето и усъвършенстването на материалите. Очаква се такъв поетапен подход да спомогне за преодоляване на потенциалните пречки пред тяхното популяризиране, разпространение и използване в ИИ общността, включително по въпроси, свързани с езика, съдържанието, контекста и наличността, които биха могли да им се отразят негативно на световно равнище, особено в Глобалния Юг.

3.2. Проект 2 — Дейности за образование и изграждане на капацитет за бъдещи специалисти в областта на ИИ

3.2.1. Цел на проекта

Целта на проект 2 е да се подкрепи интегрирането на проблема с отклоняването на граждански научни изследвания в областта на ИИ и злоупотребата с тях от безотговорни участници в образованието на бъдещите поколения специалисти в областта на ИИ. В дългосрочен план това ще гарантира, че студентите в областта на науките, технологиите, инженерството и математиката (НТИМ), които оформят бъдещето на ИИ, са наясно с отрицателното въздействие, което отклоняването на тяхната работа и злоупотребата с нея от безотговорни участници биха могли да имат върху международния мир и сигурност, и че те ще разполагат с основните инструменти, необходими за идентифициране и смекчаване на такъв риск в процеса на извършване на научните изследвания и иновациите.

В рамките на този проект ще се проведе поредица от образователни семинари и семинари за изграждане на капацитет с преподаватели и студенти в сътрудничество с избрани международни университети и представители на промишлеността. По този начин проектът има за цел да подготви дейности за изграждане на капацитет, които преподавателите и съставителите на академични програми биха могли да използват, като ги включат в обучението на бъдещите специалисти в областта на ИИ (например курсове по етика и отговорни иновации в областта на ИИ), и съображения за рисковете от

отклоняване на граждански научни изследвания и иновации в областта на ИИ и злоупотреба с тях от безотговорни участници, както и да ги свърже с по-широкия контекст на мира и сигурността. Чрез тези семинари проектът ще се стреми също така да определи мрежа от заинтересовани преподаватели, съставители на учебни програми и студенти, които да подкрепят разпространението и популяризирането на образователните материали по проекта и дейностите за изграждане на капацитет в образователната общност в областта на ИИ и в общността на специалистите в областта на ИИ. Този компонент за изграждане на мрежи има за цел да гарантира устойчивостта на проектите отвъд непосредственото им времетраене и да позволи изграждането на по-силни връзки в подкрепа на гражданската техническа ангажираност за постигане на по-големи цели за мир, сигурност, разоръжаване и контрол върху оръжията.

3.2.2. Описание на проекта

В рамките на този проект ще се проведат поредица от образователни семинари и семинари за изграждане на капацитет с преподаватели и студенти от избрани университети от цял свят. Те ще се състоят от комбинация от лекции и интерактивни дейности, които ще предоставят на преподавателите и студентите възможности да обмислят как гражданските научни изследвания и иновации в областта на ИИ биха могли да бъдат отклонени и използвани по начини, които могат да представляват риск за международния мир и сигурност, и как тези рискове могат да бъдат идентифицирани, предотвратени или смекчени в процеса на научни изследвания и иновации или чрез други процеси на управление. Тези дейности ще се основават на предходната по-маломасщабна пилотна работа, извършена от UNODA, която експериментира с методи за ангажиране и повишаване на осведомеността на студентите в областта на НТИМ към значението на отчитането на по-широкото въздействие на тяхната работа, както и на взаимодействието с експертния опит извън основните им области. По-конкретно, те ще се състоят от:

- (a) регионални семинари за изграждане на капацитет за преподаватели и студенти (4) — в рамките на регионалните семинари ще се провеждат и насърчават дейности, които преподавателите могат да използват за изграждане на капацитет на студентите в областта на НТИМ за отговорни иновации в областта на ИИ, със специален акцент върху начините за оценка и смекчаване на рисковете от отклоняване на граждански ИИ технологии и злоупотреба с тях от безотговорни участници. Всеки семинар ще бъде организиран с университет със седалище в ЕС и водещ университет от различен регион на света, като по този начин винаги ще се свързват различни групи участници от ЕС с различни групи участници извън ЕС. След това семинарите ще обхванат Латинска Америка и Карибския басейн, Северна Америка, Африка, Азия и Тихоокеанския басейн. Това ще позволи участието на студенти (по магистърски и докторски програми) от цял свят, включително от Глобалния Юг. Семинарът ще се провежда основно на английски език, но когато е осъществимо, на участниците ще бъде предоставена възможност да участват в дейности, основани на други езикови групи; и
- (b) международен семинар относно устойчивото изграждане на капацитет (1) — семинарът ще се основава на поуците, извлечени от регионалните семинари, и ще

улесни обмена на информация и опит между преподавателите и избрани студенти от университетите, участващи в проекта. На семинара ще бъдат обсъдени начините за усъвършенстване на дейностите и инструментите, разработени в хода на проекта, и тяхното разпространение извън групата на участващите университети. Ще бъдат обсъдени и начините за подпомагане на взаимодействието на студентите за отговорен ИИ, така че да се преодолеят рисковете от отклоняване и злоупотреба за международния мир и сигурност, след като те се влеят на пазара на труда.

Мрежите и присъствието на SIPRI и UNODA в Африка, Азиатско-тихоокеанския регион, Европа, Северна и Южна Америка ще бъдат използвани за улесняване и подпомагане на определени аспекти на дейностите по целесъобразност.

3.2.3. Очаквани резултати от проекта

Очаква се проектът да създаде модели за изграждане на капацитет и дейности за взаимодействие, които преподавателите и съставителите на академични програми биха могли да възпроизведат, за да повишат осведомеността на бъдещите специалисти в областта на ИИ към проблемите с отклоняването на граждански ИИ и злоупотребата с него от безотговорни участници и как те могат да спомогнат за смекчаване на тези проблеми чрез процеси за отговорни иновации. След приключване на дейностите по проекта от участниците (преподаватели, но също така и студенти в областта на НТИМ) ще се очаква да могат да използват и насърчават инструменти, методи и концепции за отговорни иновации за идентифициране и смекчаване на рисковете от отклоняване и злоупотреба при разработването и разпространението на граждански научни изследвания и иновации в областта на ИИ.

Очаква се също така дейностите по проекта да изградят мрежа от преподаватели, съставители на учебни програми и студенти, които не само ще популяризират дейностите по проекта в рамките на образователните и професионалните общности в областта на ИИ (например по време на конференции на Обществото за изчислителен интелект към IEEE), но също така ще бъдат на разположение, за да допринесат с технически капацитет за ръководени от държавите международни процеси на управление (например процеса по Конвенцията за конвенционалните оръжия относно нововъзникващите технологии в областта на смъртоносните автономни оръжейни системи).

Краткосрочната и дългосрочната стойност на тези дейности ще бъде доказана чрез проучвания преди и след тяхното изпълнение.

3.3. Проект 3 — Улесняване на по-дългосрочното устойчиво разработване, разпространение и въздействие на отговорните иновации в областта на ИИ за целите на мира и сигурността

3.3.1. Цел на проекта

Целта на проект 3 е да се улесни по-дългосрочното устойчиво разработване, разпространение и въздействие на отговорните иновации в областта на ИИ като средство за намаляване на рисковете, които отклоняването на граждански научни изследвания и

иновации в областта на ИИ и злоупотребата с тях могат да създадат за мира и сигурността. Намерението е той да бъде осъществен чрез кръгли маси с представители на сектора на ИИ, диалози с участието на множество заинтересовани страни, изготвяне на публичен доклад и целенасочени дейности за разпространение. Проектът има за цел да гарантира, че генерираната работа, по-специално дейностите в областта на образованието, изграждането на капацитет и взаимодействието, достигат до ИИ общността и ѝ оказват въздействие като цяло, на всички равнища (от студенти до инженери и други специалисти в областта на ИИ) и отвъд географските, секторните и други граници. За да се увеличи възможността за широко и дълбоко въздействие, от съществено значение е да се сътрудничи с професионалните организации в това пространство, като например IEEE, и да се осъществява многостранно взаимодействие с академичните среди, промишлеността и други сектори. Подобни усилия ще дадат възможност на заинтересовани представители от различни ИИ общности да поемат отговорност за проблема и да представят собствени възгледи за устойчивото реализиране и насърчаване на усилията за смекчаване на риска в рамките на световната ИИ общност. За дългосрочната значимост на проекта за държавите, междуправителствените организации и др. е важно специалистите в областта на ИИ да могат да се учат от правителствени експерти, ангажирани с намаляването на риска в контекста на разоръжаването, контрола върху оръжията и неразпространението на оръжия, и да си взаимодействат с тях. От решаващо значение за устойчивостта е също така да се гарантира, че информацията, получена чрез дейностите в областта на взаимодействието, се анализира, консолидира и разпространява по подходящ начин.

3.3.2. Описание на проекта

Този проект се състои от следните ключови направления:

- (a) диалози с участието на множество заинтересовани страни на тема „Отговорни иновации в областта на ИИ за мир и сигурност“ (до 9) — тази поредица от виртуални срещи диалози ще събере експерти от академичните и научноизследователските среди, частния сектор и традиционния контрол върху оръжията от ЕС и извън него, за да обсъдят:
- i. технологичните тенденции, които могат да породят рискове от разпространение, отклоняване и злоупотреба с последици за международния мир и сигурност;
 - ii. как да се допринесе за намаляването на риска чрез процеси, методи и средства за отговорни иновации, както и възможностите и предизвикателствата, свързани с диалога и обмяна на знания между общностите на заинтересованите страни, включително тези, които работят в други сектори, като например биологичната и химическата промишленост; и
 - iii. потенциалната стойност, цел и формат на една самоподдържаща се мрежа от експерти и дейности за диалог. Експертната група ще заседава няколко пъти годишно и ще работи за организирането на две публични прояви за по-широката общност.

От деветте виртуални срещи две ще бъдат отворени за обществеността, за да се улесни консултирането в по-широк кръг.

(b) кръгли маси с частния сектор (до 6) — тази поредица от виртуални кръгли маси ще постави началото на диалог с участници, работещи с процеси за отговорни иновации в областта на ИИ в частния сектор (напр. партньорство в областта на ИИ), за това как те могат да допринесат за свеждане до минимум на рисковете от отклоняване на граждански технологии с ИИ и злоупотреба с тях от безотговорни участници, както и ще проучи възможните стимули за това в рамките на развитието на частния сектор. Темите ще включват:

- i. значението на контекста на международната сигурност и разоръжаването за частния сектор;
- ii. правната среда, в която съществуват разработването, внедряването и експлоатацията на ИИ по света;
- iii. как да се надгражда върху механизми за оценка на риска и други мерки, които са част от или биха могли да бъдат интегрирани в процеси за отговорни иновации и корпоративни програми за съответствие; и
- iv. поуки, които трябва да се извлекат от други отрасли, процеси и нормативни уредби, свързани с контрола върху оръжията (напр. биологичната и химическата промишленост).

(c) доклад относно гледните точки на ИИ общността относно контрола върху оръжията и намаляването на риска в областта на ИИ, насочен към ИИ общностите и тези за контрол върху оръжията (1) — изготвянето на този доклад ще обхване и консолидира констатациите и препоръките на проекта в единен референтен документ, насочен както към отговорната за гражданските приложения на ИИ общност, така и към общността за контрол върху оръжията. В доклада ще се разглеждат начините за идентифициране, оценка и преодоляване на рисковете за международния мир и сигурност, свързани с отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях.

(d) прояви за разпространение, насочени към консултации и взаимодействие с ИИ общността и общността за контрол върху оръжията (предстои да се уточни) — екипът ще търси възможности за представяне на работата и на констатациите и препоръките, свързани с нея, по време на целия проект. Форматът на проявите и съдържанието на презентациите ще бъдат съобразени с нуждите на целевите групи. Те могат да включват заседания на работна група „Неразпространение и износ на оръжия“ (CONOP), общото събрание на Европейския алианс за изкуствен интелект; групата от правителствени експерти по нововъзникващи технологии в областта на смъртоносните автономни оръжейни системи; междуведомствената работна група по ИИ (IAWG-AI); инициативата на Международния съюз по далекосъобщения (МСД) „ИИ за добро“; годишния диалог по въпросите на иновациите в рамките на Института на ООН по разоръжаването (ЮНИДИР); и Института на електроинженерите и инженерите по електроника. Екипът ще се

стреми също така да работи двустранно със съответните заинтересовани страни от държавния сектор, академичните среди и частния сектор.

3.3.3. Очаквани резултати от проекта

Очаква се този проект да положи основите за устойчивото развитие, разпространение и въздействие на отговорните иновации в процесите на ИИ, насочени към рисковете от разпространение, отклоняване и злоупотреба и последиците от тях за мира и сигурността, след непосредствения край на срока на действие на решението на Съвета.

Очаква се диалогът с множество заинтересовани страни да предостави модел за обмен на информация и сътрудничество за намаляване на риска не само в рамките на световната ИИ общност, но и между отговорната за гражданските приложения на ИИ общност и общностите в областта на разоръжаването, контрола върху оръжията и неразпространението на оръжия. Такъв модел би могъл да се използва за запознаване на създателите на политики с ключовия технологичен и научен напредък, свързан с отговорните иновации в областта на ИИ, както и за запознаване на техническата аудитория със средата, в която понастоящем действат създателите на политики. Очаква се проектът да улесни устойчивите отношения и взаимодействието между заинтересованите участници в рамките на тези различни общности и между тях. Очаква се тези разнородни мрежови ефекти да дадат възможност за по-голямо развитие и широко разпространение на отговорните иновации в областта на ИИ за мир и сигурност след изтичане на срока на проекта.

Очаква се диалогът с частния сектор да даде възможност за по-голямо и по-задълбочено участие на частния сектор в областта на ИИ в идентифицирането, предотвратяването и намаляването на рисковете за мира и сигурността, произтичащи от отклоняването и злоупотребата с граждански научни изследвания и иновации в областта на ИИ. Очаква се проектът да допринесе за по-доброто разбиране и ангажираност с проблемите, които се стреми да разреши, от страна на ключовите участници в процесите от частния сектор. Освен това той има за цел да улесни (по-широкото) възприемане и прилагане на процеси, методи и средства за отговорни иновации в съществуващите механизми и процедури за управление на корпоративния риск.

Очаква се също така диалогът с участието на множество заинтересовани страни и кръглата маса в частния сектор да дадат представа по редица съществени въпроси, включително а) как методите и средствата за отговорни иновации могат да бъдат допълнително усъвършенствани и използвани за идентифициране, предотвратяване и намаляване на рисковете, породени от отклоняването на граждански научни изследвания и иновации в областта на ИИ и злоупотребата с тях; б) как научните изследвания и иновациите в областта на ИИ могат да се използват положително в подкрепа на целите на мира и сигурността (напр. приложения за ранно предупреждение за конфликти и хуманитарна помощ); и в) как да се улесни по-активен диалог и обмен на информация относно усилията за намаляване на риска, предприети в отговорната за гражданските приложения на ИИ общност (т.е. различни инициативи, ръководени от IEEE) и

инициативите, които вече се провеждат в общността за разоръжаване, контрол на оръжията и неразпространение на оръжия на междуправителствено равнище.

Докладът и дейностите по разпространение ще анализират, консолидират и разпространят информацията, получена по проекти 1, 2 и 3, и по този начин ще подкрепят популяризирането на констатациите и препоръките от дейностите по проектите в рамките на световната ИИ общност, а така също и в рамките на политическата общност. Очаква се също така те да спомогнат за осигуряване на устойчивост на въздействието след изтичането на срока на проектите.

4. ПРОДЪЛЖИТЕЛНОСТ

Общата планирана продължителност на проектите е 36 месеца.