



Euroopan unionin
neuvosto

Bryssel, 26. lokakuuta 2022
(OR. en)

14095/22
ADD 1

LIMITE

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

EHDOTUS

Lähettiläjä:	Unionin ulkoasioiden ja turvallisuuspolitiikan korkea edustaja, allekirjoittajana pääsihteeri Stefano SANNINO
Saapunut:	26. lokakuuta 2022
Vastaanottaja:	Thérèse BLANCHET, Euroopan unionin neuvoston pääsihteeri
Asia:	Unionin ulkoasioiden ja turvallisuuspolitiikan korkean edustajan neuvostolle esittämä ehdotus neuvoston päätökseksi vastuullisen innovoinnin edistämistä tekoälyn alalla rauhaa ja turvallisuutta varten koskevan hankkeen täytäntöönpanon tukemisesta

Valtuuskunnille toimitetaan oheisena asiakirja HR(2022) 238.

Liite: HR(2022) 238

EUROOPAN ULKOSUHDEHALLINTO



HANKEASIAKIRJA

Vastuullisen tekoälyinnovoinnin edistäminen rauhaa ja turvallisuutta varten

HR(2022)238

1. TAUSTAA

Viimeaikainen kehitys tekoälyalalla on avannut uusia mahdollisuuksia rauhan ja turvallisuuden tukemista ja säilyttämistä varten, esimerkiksi teknologisilla parannuksilla sellaisilla aloilla, kuten konfliktien varhaisvaroitusta sekä aseiden ja niiden viennin valvonnan todentaminen. Toisaalta nämä kehitysaskleet ovat antaneet mahdollisuuden uusille tavoille luoda tai kiihdyttää jännitteitä, konflikteja ja turvattomuutta valtioissa tai niiden välillä. Riskit, joita aiheutuu tietyistä tekoälyn soveltamistavoista, kuten tappavista autonomisista asejärjestelmistä, ovat nousseet merkittäväksi huolenaiheeksi asevalvontayhteisössä. Yksi riskien osa-alue, johon on kiinnitettävä enemmän huomiota ja jonka käsittelemiseen nykyiset asevalvonta- ja diplomaattiset toimet saattavat olla riittämättömiä, on vastuuttomien toimijoiden, myös vihamielisten valtiosta riippumattomien toimijoiden, harjoittama tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö ja väärinkäyttö, kuten generatiivisten kilpailevien verkkojen (GAN) väärinkäyttö syvävääreännösten (deepfake) tuottamiseen disinformaatiokampanjoita varten.

Tekoäly on mahdollistava teknologia, jolla on merkittävä yleiskäyttöpotentiaali. Tiedetyt toimijat voisivat (suhteellisen helposti) päästä käsiksi siviilisovelluksia varten kehitettyyn tekoälytutkimukseen ja -innovointeihin ja ottaa sen omiin haitallisiin tai häiriötä aiheuttaviin käyttötarkoituksiinsa, joilla voi olla vaikutuksia kansainväliseen rauhaan ja turvallisuuteen. Siviiliteknologian tarkoituksenvastainen käyttö ja väärinkäyttö eivät ole uusia ilmiöitä eikä tätä tapahdu vain tekoälyalalla. Siihen liittyvällä robotiikan alalla tästä on äskettäisenä osoituksena se, että Da'esh/Isis aseisti ja otti käyttöön virkistyskäyttöön tarkoitettuja miehittämättömiä ilma-aluksia Syyriassa. Tekoälyn tapauksessa ongelmaa mutkistavat kuitenkin monet tekijät: tekoälyalgoritmien ja -datan aineeton ja nopeasti muuttuva luonne, minkä vuoksi niiden siirtoa tai levitystä on vaikea valvoa; yksityisen sektorin johtava rooli tutkimus-, kehitys- ja innovointiekosysteemeissä ja sen tarve suojella omistusoikeuden alaisia algoritmeja; sekä maailmanlaajuinen saatavuus sellaiselle inhimilliselle asiantuntemukselle ja aineellisille resursseille, joiden avulla on mahdollista ottaa tekoälyteknologioita uusiin käyttötarkoituksiin.

Tekoälyn parissa siviilialalla työskentelevät henkilöt ovat samaan aikaan liian usein tietämättömiä mahdollisista vaikutuksista, joita heidän työnsä tarkoituksenvastaisella käytöllä ja väärinkäytöllä voisi olla kansainväliselle rauhalle ja turvallisuudelle, tai he epäröivät osallistua käynnissä oleviin keskusteluihin tekoälystä aiheutuvista riskeistä asevalvonnassa ja asesulkualalla.

On tarpeen tukea tekoälyalan siviiliyhteisön laajempaa osallistumista sellaisten rauhaan ja turvallisuuteen liittyvien riskien ymmärtämiseen ja lieventämiseen, joita aiheutuu siitä, että vastuuttomat toimijat käyttävät siviilialan tekoälyteknologiaa tarkoituksenvastaisesti tai väärinkäyttävät sitä. Tukholman kansainvälisen rauhantutkimuslaitoksen (SIPRI) ja Yhdistyneiden kansakuntien aseriisuntatoimiston (UNODA) osalta tämä voitaisiin saavuttaa edistämällä vastuullista innovointia maailmanlaajuisessa tekoälyalan siviiliyhteisössä. SIPRI:n ja UNODAn aiempi työ on osoittanut, että vastuullinen innovointi itsehallinnollisena mekanismina voisi tarjota maailmanlaajuiselle tekoälyalan siviiliyhteisölle käytännön välineitä ja menetelmiä sellaisten riskien tunnistamiseksi, joita tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö ja väärinkäyttö voisi aiheuttaa rauhalle ja turvallisuudelle, sekä auttaa ehkäisemään ja lieventämään näitä riskejä. SIPRI:n ja UNODAn työssä yksilöitiin myös menetelmiä ja useita käynnissä olevia vastuulliseen tekoälyyn keskittyviä siviililähtöisiä aloitteita, joiden pohjalta tekoälyalan siviiliyhteisöä voitaisiin valistaa asevalvontaa ja asesulkua koskevista kysymyksistä, asiantuntijakeskusteluista ja valtion kannoista tekoälyn vastuulliseen kehittämiseen, levittämiseen ja käyttöön sekä puolustusalan vastuuta koskevasta työstä saaduista kokemuksista.¹ Keskeistä on, että aiemmassa työssä on selkeästi todettu, että yhteydet luonnontieteiden, teknologian, insinööritieteiden ja matematiikan (STEM-aineet) opiskelijoihin, jotka käyttävät tekoälyä opetusympäristössä, ovat keskeisellä sijalla kaikissa tehokkaissa vastuullisissa innovointitoimissa.

2. TAVOITTEET

Hankkeiden tavoitteena on tukea tekoälyalan siviiliyhteisön laajempaa osallistumista niiden riskien lieventämiseen, joita vastuuttomien toimijoiden harjoittama tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa kansainväliselle rauhalle ja turvallisuudelle. Tavoitteena on ensinnäkin saada parempi käsitys siitä, miten tekoälyalan tutkimuksen ja -innovoinnin kehittämistä ja levittämistä koskevat päätökset voivat vaikuttaa tarkoituksenvastaisesta käyttöä ja väärinkäyttöä koskeviin riskeihin sekä luoda rauhaa ja turvallisuutta koskevia riskejä tai mahdollisuuksia. Toiseksi pyritään edistämään vastuullisia innovointiprosesseja, -menetelmiä ja -välineitä, jotka voivat auttaa varmistamaan siviili-innovaatioiden rauhanomaisen soveltamisen ja tekoälyä koskevan tietämyksen vastuullisen levittämisen. Tätä varten hankkeilla tuetaan valmiuksien kehittämiseen, tutkimukseen ja osallistamiseen tähtääviä toimia, joilla *i)* parannetaan maailmanlaajuisen tekoälyalan siviiliyhteisön valmiuksia ottaa huomioon ja käsitellä vastuullisten innovointiprosessien avulla rauhaa ja turvallisuutta koskevia riskejä, joita aiheutuu vastuuttomien toimijoiden harjoittamasta siviilialan tekoälyn tarkoituksenvastaisesta käytöstä tai väärinkäytöstä; ja *ii)* lujitetaan vastuulliseen tekoälyyn liittyvien riskien siviilitason lieventämistoimien ja hallitustenvälisellä tasolla jo käynnissä olevien aseriisunta-, asevalvonta- ja asesulkutoimien välistä yhteyttä. Ratkaisevan tärkeää on, että hankkeilla ei pyritä luomaan uusia standardeja, periaatteita tai sääntelyä taikka muutoin toimimaan valtioiden toimivaltaan kuuluvilla alueilla.

¹ Menetelmiä ovat esimerkiksi kansainvälisen sähkö- ja elektroniikkainsinöörien järjestön (IEEE) suosittelemat käytännöt, joilla arvioidaan autonomisten ja älykkäiden järjestelmien vaikutusta ihmisten hyvinvointiin (IEEE-standardi 7010-2020), ja tekoälyä käsittelevän korkean tason asiantuntijaryhmän luotettavan tekoälyn arviointilista. Aloitteita ovat esimerkiksi autonomisten ja älykkäiden järjestelmien eettisiä periaatteita koskeva IEEE:n maailmanlaajuinen hanke; tekoälyä koskeva kumppanuus; tekoälyä koskeva maailmanlaajuinen kumppanuus.

Sen sijaan niiden tarkoituksena on kehittää vastuullisia siviilialan innovointitoimia, jotta voidaan ottaa huomioon rauhaan ja turvallisuuteen liittyvät riskit, joita aiheutuu vastuuttomien toimijoiden harjoittamasta siviilialan tekoälyn tarkoituksenvastaisesta käytöstä tai väärinkäytöstä, sekä tarjota koulutusta nykyisistä asiaankuuluvista hallitustenvälisistä toimista.

Jotta hankkeet tavoittaisivat tekoälyalan siviiliyhteisön ja vaikuttaisivat tehokkaasti, niissä sovelletaan kolmiosaista lähestymistapaa, jolla pyritään seuraaviin:

- a) yhteistyö kouluttajien kanssa – tehdään yhteistyötä valikoitujen opettajien ja akateemisten opetussuunnitelmien kehittäjien kanssa liittyen tulevien tekoälyalan ammattilaisten koulutukseen (esim. tekoälyn etiikkaa ja vastuullista innovointia käsittelevät kurssit) sellaisten opetusmateriaalien kehittämiseksi ja edistämiseksi, joita voidaan käyttää ottamaan huomioon rauhaa ja turvallisuutta koskevat riskit, joita aiheutuu siitä, että vastuuttomat toimijat käyttävät tarkoituksenvastaisesti ja väärinkäyttävät tekoälyalan siviilitutkimusta ja -innovointia;
- b) yhteistyö opiskelijoiden kanssa – esitellään valikoiduille ja eri puolilta maailmaa tuleville luonnontieteiden, teknologian, insinööritieteiden ja matematiikan (STEM-aineet) opiskelijoille, miten vastuuttomien toimijoiden harjoittamasta tekoälyalan siviilikehitystyön tarkoituksenvastaisesta käytöstä ja väärinkäytöstä aiheutuvat rauhaa ja turvallisuutta koskevat riskit voidaan tunnistaa ja miten niitä voidaan estää tai lieventää tutkimus- ja innovointiprosessissa tai muissa hallinnointiprosesseissa; ja
- c) yhteistyö tekoälyteollisuuden kanssa – tehdään yhteistyötä ammatillisten järjestöjen ja standardointielinten, kuten kansainvälisen sähkö- ja elektroniikkainsinöörien järjestön (IEEE) kanssa, jotta voidaan i) jakaa räätälöityjä opetusmateriaaleja ja osallistamiseen tähtääviä toimia teknisen alan ammattilaisille; ii) tukea tekoälyn myönteisiä käyttötarkoituksia rauhan ja turvallisuuden edistämiseksi; ja iii) helpottaa tiedemaailman, yksityisen sektorin ja hallituksen asiantuntijoiden välistä vuoropuhelua ja tiedonvaihtoa liittyen sen riskin lieventämiseen, että vastuuttomat toimijat käyttäisivät tarkoituksenvastaisesti ja väärinkäyttäisivät tekoälyalan siviilitutkimusta ja -innovointia.

Tällainen lähestymistapa mahdollistaa sen, että hankkeet tavoittavat tekoälyalan yhteisön kaikilla tasoilla, eli nykyisten toimijoiden lisäksi myös tulevat sukupolvet. Se mahdollistaa myös akateemisten, teollisten ja muiden alojen osallistumisen ja tukee tulevien toimien kestävyttä perustamalla rajatylittäviä verkostoja.

Hankkeissa pyritään myös hyödyntämään SIPRI:n ja UNODAn vaikutusvaltaa ja kokemusta, jotta tekoälyalan yhteisöön, ei pelkästään EU:n sidosryhmiin, voidaan vaikuttaa maailmanlaajuisesti. SIPRI ja UNODA ovat ainutlaatuisessa asemassa luomaan yhteyksiä tekoälyalan toimijoiden yhteistyöhön Afrikan, Aasian ja Tyynenmeren alueen, Euroopan sekä Pohjois- ja Etelä-Amerikan välillä sekä edistämään tätä yhteistyötä. Molemmilla yhteisöillä on kokemusta myös muilla tieteen ja teknologian aloilla, kuten bioteknologia, joilla on samanlaisia kaksoiskäyttöä ja leviämistä koskevia haasteita. Hankkeissa pyritään myös hyödyntämään Euroopan unionissa vallitsevia olosuhteita, joita ovat muun muassa a) vastuullista tekoälyä koskevat kehittyneet monen sidosryhmän prosessit, b) EU:n aseriisunta-, asevalvonta- ja asesulkukysymyksiin sitoutumisen ja niihin liittyvän asiantuntemuksen korkea taso; c) EU:n akateemisten organisaatioiden, tutkimusorganisaatioiden ja yksityisen sektorin organisaatioiden ja muiden alueiden, erityisesti globaalin etelän, välisten yhteyksien moninaisuus, joka on myös tärkeä osa osallistamista; ja d) opiskelijoiden, kouluttajien ja insinöörien kansalaisuuksien moninaisuus yliopistoissa, tutkimuslaitoksissa ja yksityisellä sektorilla.

Osallistaminen on yksi keskeisistä näkökohdista hanketoimien toteuttamisessa. Jotta tekoälyalan yhteisöä voitaisiin tukea tehokkaasti, hankkeissa todetaan, että tekoälyalan yhteisö koostuu monenlaisista toimijoista ja erityisesti että sukupuoli on erittäin tärkeä tekijä.

- a) Tästä syystä sukupuolinäkökulma otetaan huomioon YK:n järjestelmän laajuisissa sukupuolten tasa-arvon valtavirtaistamista ja tasa-arvoa koskevissa strategioissa. Naisten osallistumista kaikkiin hankkeen toimiin kannustetaan ja sitä edellytetään, ja
- b) vammaisten henkilöiden ja tarpeiden kohtuullista mukauttamista edellyttävien henkilöiden osallistamista toteutetaan kokonaisvaltaisesti. Tähän sisältyy vammaisten henkilöiden osallistumisen esteiden poistaminen. Lisäksi varmistetaan, että toteutetaan toimia vammaisten henkilöiden kannalta tärkeiden näkemysten ja kokemusten huomioon ottamiseksi ja niiden edustuksen edistämiseksi.

3. HANKKEET

Jäljempänä kuvattujen kolmen hankkeen on tarkoitus olla toisiaan täydentäviä ja toisiaan tukevia, ja niiden osia toteutetaan 36 kuukauden ajanjaksolla.

Hanke 1 – Koulutusta ja valmiuksien kehittämistä koskevan materiaalin tuottaminen tekoälyalan siviiliyhteisöä varten

3.1.Hankkeen tarkoitus

Hankkeessa nro 1 keskitytään tarjoamaan tekoälyalan siviilitoimijoille tietoa ja keinoja, joiden avulla voidaan arvioida ja lieventää riskejä, joita vastuuttomien toimijoiden harjoittama tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa liittyen kansainväliseen rauhaan ja turvallisuuteen. Sen tavoitteena on tuottaa koulutusta ja valmiuksien kehittämistä koskevaa materiaalia, jonka avulla tekoälyalan ammattilaiset kaikilta alueilta, tasoilta ja aloilta (mukaan lukien tekoälyyn keskittyvät kouluttajat, opetussuunnitelmien kehittäjät, STEM-aineiden opiskelijat ja tekoälyinsinöörit sekä tiedemaailman ja yksityisen sektorin tutkijat) saavat tarvittavat tiedot ja välineet, jotta on mahdollista

- a) ymmärtää, miten tekoälyalan siviilitutkimusta ja -innovointia voitaisiin käyttää tarkoituksenvastaisesti ja väärinkäyttää tavoilla, jotka voisivat aiheuttaa kansainvälistä rauhaa ja turvallisuutta koskevia riskejä, ja miten tutkimuksen ja innovoinnin kehittämistä ja levittämistä koskevat päätökset voivat lisätä tai vähentää tarkoituksenvastaisen käytön ja väärinkäytön riskiä;
- b) ymmärtää aseriisunta-, asevalvonta- ja asesulkuyhteisön jo toteuttamia toimia, joilla pyritään lieventämään siviilitutkimuksen ja -innovoinnin tarkoituksenvastaisen käytön ja väärinkäytön riskiä; ja
- c) harjoittaa vastuullista innovointia tavalla, joka vähentää tarkoituksenvastaisen käytön ja väärinkäytön riskiä tutkimuksen ja innovoinnin kehittämisessä ja levittämisessä.

3.1.1. Hankkeen kuvaus

Hankkeessa tuotetaan kolme erillistä koulutusta ja valmiuksien kehittämistä koskevaa materiaalityypistä.

- a) *Käsikirja (1)* – Käsikirjaan kootaan perustiedot ja keinot, joiden avulla tekoälyalan toimijat voivat arvioida ja lieventää tutkimus- ja innovointiprosessissa riskejä siitä, että vastuuttomat toimijat käyttävät tarkoituksenvastaisesti ja väärinkäyttävät tekoälyalan siviiliteknologiaa. Siinä käsitellään sitä, miksi ja miten tutkimuksen ja innovoinnin kehittämistä ja levittämistä koskevat päätökset voivat vaikuttaa tarkoituksenvastaisen käytön ja väärinkäytön riskeihin ja vuorostaan luoda rauhaan ja turvallisuuteen liittyviä riskejä tai mahdollisuuksia.

Siinä esitellään myös asiakoskeva kansainvälinen oikeus ja vientivalvontavelvoitteet sekä sotilasalalla ja aseriisunta-, asevalvonta- ja asesulkupiireissä käsiteltävät turvallisuusnäkökohdat. Lisäksi siinä esitellään esimerkkiprosesseja ja -välineitä vastuullisen innovoinnin harjoittamiseksi, kuten teknologian vaikutustenarviointimenetelmiä ja riskinarviointimalleja.

- b) *Podcast-sarja (~10)* – Näiden podcast-jaksojen avulla tekoälyalan toimijat saavat tietoa siitä, miksi ja miten vastuulliset tekoälyn innovointiprosessit voivat tukea kansainvälistä rauhaa ja turvallisuutta lieventämällä riskejä, joita aiheutuu vastuuttomien toimijoiden harjoittamasta tarkoituksenvastaisesta käytöstä ja väärinkäytöstä. Sarjassa tarkastellaan tärkeitä teemoja (esim. kaksi- ja yleiskäyttöä koskevan tutkimuksen ja innovoinnin tarkoituksenvastainen käyttö ja väärinkäyttö; humanitaariset, strategiset ja poliittiset haasteet, jotka liittyvät tekoälyalan siviilitutkimuksen ja -innovoinnin mahdolliseen väärinkäyttöön; aseriisunta-, asevalvonta- ja asesulkupiirien kohtaamat haasteet pyrkimyksissä lieventää riskejä; vastuullisen innovoinnin toteuttaminen riskinarvioinnin avulla; vientivalvonnan noudattaminen; riskien suunnitelmallinen vähentäminen; vastuullinen julkaiseminen; asiakkaiden tunteminen; ja suunnitteluharjoituksista saatava kokemus). Sarja rakentuu haastattelujen ympärille, joita hankeryhmä käy asiaankuuluvien yhteisöjen edustajien kanssa.
- c) *Blogisarja (9–10)* – Ryhmä kehittää räätälöidyn blogisarjan, jonka tarkoituksena on tuoda esiin toimia, joilla pyritään luomaan yhteyksiä ”vastuulliseen tekoälyyn” liittyvien siviilikeskeisten yhteisöjen ja asevalvonta- ja asesulkuyhteisöjen välisten rajojen yli. Blogisarja tarjoaa foorumin, jolla levitetään näkemyksiä, ideoita ja ratkaisuja, jotka koskevat siviilitekoälyn tarkoituksenvastaiseen käyttöön ja väärinkäyttöön tutkimus- ja innovointiprosessissa liittyvien riskien tunnistamista ja niihin puuttumista. Blogissa pyritään tuomaan esiin tekoälyalalla vallitsevien ajatusten ja näkökulmien moninaisuutta.

Tätä aineistoa levitetään julkisesti täytäntöönpanevien toimijoiden verkkosivustoilla, niiden sosiaalisen median profiilien kautta ja suorassa viestinnässä asiaankuuluvien akateemisten tahojen, siviilitekoälyn ammattijärjestöjen ja muiden asiaankuuluvien ryhmien kanssa.

3.1.2. Hankkeelta odotetut tulokset

Tällä hankkeella on tarkoitus luoda uusia materiaaleja, joiden avulla tekoälyalan siviilitoimijat voivat saada tietoa a) siitä, miten tekoälyalan siviilitutkimusta ja -innovointia voitaisiin käyttää tarkoituksenvastaisesti ja väärinkäyttää tavoilla, jotka voivat aiheuttaa kansainväliseen rauhaan ja turvallisuuteen liittyviä riskejä, b) miten aseriisunta-, asevalvonta- ja asesulkuyhteisö käsittelee tällaisia riskejä ja c) miten tekoälyalan toimijat voisivat edelleen edistää tällaisten riskien lieventämistä vastuullisten innovointiprosessien avulla.

Tällä pyritään edistämään tekoälyalan siviiliyhteisön laajempaa osallistumista niiden riskien lieventämiseen, joita siviilitekoälyn tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa kansainväliselle rauhalle ja turvallisuudelle. Lisäksi pyritään parantamaan teknisten toimijoiden valmiuksia osallistua aseriisunta-, asevalvonta- ja asesulkuyhteisön asiaankuuluviin prosesseihin sekä tukemaan sellaisten uusien yleisöjen osallistamista, jotka eivät perinteisesti ole osallisena aseriisunta- ja asesulkualan koulutustoimissa.

Materiaalin odotetaan myös tukevan muiden hankkeiden täytäntöönpanoa, ja sitä käytetään perustana hankkeen nro 2 koulutusta ja valmiuksien kehittämistä koskeville toimille sekä hankkeen nro 3 vuoropuhelu- ja osallistamistoimille. Näiden toimien odotetaan puolestaan edistävän materiaalin tuotantoa ja kehittämistä. Tällaisen iteratiivisen lähestymistavan odotetaan auttavan poistamaan materiaalin edistämisen, levittämisen ja käytön mahdollisia esteitä tekoälyalan yhteisössä, mukaan lukien kieleen, sisältöön, asiayhteyteen ja saatavuuteen liittyvät kysymykset, jotka voisivat haitata sen vaikutusta maailmanlaajuisella tasolla, erityisesti globaalissa etelässä.

3.2. Hanke 2 – Koulutusta ja valmiuksien kehittämistä koskevat toimet tulevia tekoälyalan ammattilaisia varten

3.2.1. Hankkeen tarkoitus

Hankkeen nro 2 tarkoituksena on tukea sitä, että osaksi tulevien sukupolvien tekoälyalan ammattilaisten koulutusta integroidaan ongelma siitä, että vastuuttomat toimijat voivat käyttää tarkoituksenvastaisesti ja väärinkäyttää tekoälyalan siviiliteknologiaa. Näin varmistetaan pitkällä aikavälillä, että tekoälyn tulevaisuutta muokkaavat luonnontieteiden, teknologian, insinööritieteiden ja matematiikan (STEM-aineiden) opiskelijat ovat tietoisia kielteisistä vaikutuksista, joita heidän työnsä tarkoituksenvastaisella käytöllä ja väärinkäytöllä, jota vastuuttomat toimijat voivat harjoittaa, voi olla kansainväliseen rauhaan ja turvallisuuteen, ja että heillä on tarvittavat perusvälineet tällaisten riskien tunnistamiseksi ja lieventämiseksi tutkimus- ja innovointiprosessissa.

Hankkeessa järjestetään koulutus- ja valmiuksienkehittämistyöpajoja kouluttajille ja opiskelijoille yhteistyössä valittujen kansainvälisten yliopistojen ja teollisuuden toimijoiden kanssa.

Hankkeella pyritään näin ollen luomaan valmiuksienkehittämistoimia, joita kouluttajat ja akateemisten opetussuunnitelmien kehittäjät voisivat käyttää osana tulevien tekoälyalan ammattilaisten koulutusta (esim. tekoälyn etiikkaa ja vastuullista innovointia koskevat kurssit) ja pohtia riskejä, joita vastuuttomien toimijoiden harjoittama tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa, ja yhdistää ne laajempaan rauhaa ja turvallisuutta koskevaan kontekstiin. Työpajojen avulla hankkeella pyritään myös kehittämään asiasta kiinnostuneiden kouluttajien, opetussuunnitelmien kehittäjien ja opiskelijoiden verkosto, joka tukisi hankkeen opetusmateriaalin ja valmiuksienkehittämistoimien levittämistä ja edistämistä tekoälyalan koulutusyhteisössä ja ammattilaisyhteisössä. Verkostoitumisen osa-alueella pyritään varmistamaan hankkeiden kestävyys myös niiden aktiivisena olon jälkeen ja mahdollistamaan vahvempien yhteyksien luominen, jotta voidaan tukea siviilialan teknistä sitoutumista laajempien rauhaa, turvallisuutta, aseriisuntaa ja asevalvontaa koskevien tavoitteiden saavuttamiseksi.

3.2.2. Hankkeen kuvaus

Hankkeessa järjestetään koulutus- ja valmiuksienkehittämistyöpajoja ympäri maailmaa valikoitujen yliopistojen kouluttajien ja opiskelijoiden kanssa. Työpajat koostuisivat luentojen ja interaktiivisten toimien yhdistelmästä, joka tarjoaisi kouluttajille ja opiskelijoille mahdollisuuksia pohtia, miten tekoälyalan siviilitutkimusta ja -innovointia voitaisiin käyttää tarkoituksenvastaisesti ja väärinkäyttää tavoilla, jotka voivat aiheuttaa kansainväliseen rauhaan ja turvallisuuteen liittyviä riskejä, ja miten tällaisia riskejä voidaan tunnistaa, ehkäistä tai lieventää tutkimus- ja innovointiprosessissa tai muissa hallinnointiprosesseissa. Nämä toimet perustuvat UNODAn aiemmin toteuttamaan pienimuotoiseen pilottityöhön, jossa kokeiltiin menetelmiä, joilla STEM-aineiden opiskelijat saataisiin osallistumaan ja tiedostamaan, että heidän työnsä laajemmat vaikutukset ja asiantuntemuksen hyödyntäminen oman alansa ulkopuolella ovat tärkeitä. Käytännössä ne koostuisivat seuraavista:

- a) *kouluttajille ja opiskelijoille suunnatut alueelliset valmiuksienkehittämistyöpajat (4)* – Alueellisissa työpajoissa toteutetaan ja edistetään toimia, joita kouluttajat voivat hyödyntää kehittääkseen STEM-aineiden opiskelijoiden vastuullisen tekoälyn innovointiin liittyviä valmiuksia keskittyen erityisesti siihen, miten voidaan arvioida ja lieventää riskejä, joita aiheutuu siitä, että vastuuttomat toimijat käyttävät tekoälyalan siviiliteknologiaa tarkoituksenvastaisesti ja väärinkäyttävät sitä. Kukin työpaja järjestetään EU:ssa sijaitsevan yliopiston ja eri maailman alueelta peräisin olevan korkeatasoisen yliopiston kanssa. Näin EU:hun sijoittautuneiden erilaisten osallistujien joukko yhdistetään aina EU:n ulkopuolelle sijoittautuneisiin erilaisten osallistujien joukkoihin. Työpajat kattavat siten Latinalaisen Amerikan ja Karibian alueen, Pohjois-Amerikan, Afrikan sekä Aasian ja Tyynenmeren alueen. Tämä mahdollistaa sen, että (maisteri- ja tohtoritason) opiskelijat ympäri maailmaa, myös globaalista etelästä, voivat osallistua. Työpaja järjestettäisiin pääasiassa englanniksi, mutta osallistujille tarjottaisiin mahdollisuuksien mukaan tilaisuus osallistua toimintaan vaihtoehtoisten kieliryhmien perusteella; ja
- b) *kestävien valmiuksien kehittämistä käsittelevä kansainvälinen työpaja (1)* – Työpajassa hyödynnetään alueellisista työpajoista saatuja kokemuksia ja helpotetaan tietojen ja kokemusten vaihtoa kouluttajien ja hankkeeseen osallistuvista yliopistoista valikoitujen opiskelijoiden välillä.

Työpajassa keskusteltaisiin siitä, miten hankkeen aikana kehitettyjä toimia ja välineitä voitaisiin hioa ja miten niitä voitaisiin levittää osallistuvien yliopistojen ulkopuolelle. Keskustelua käytäisiin myös siitä, miten opiskelijoiden työelämään siirtymisen jälkeen voidaan tukea opiskelijoiden osallistumista liittyen vastuulliseen tekoälyyn, jolla puututaan riskeihin, joita tarkoituksenvastainen käyttö ja väärinkäyttö aiheuttaa kansainvälisen rauhan ja turvallisuuden osalta.

SIPRI:n ja UNODAn verkostoja ja läsnäoloa Afrikassa, Aasian ja Tyynenmeren alueella, Euroopassa sekä Pohjois- ja Etelä-Amerikassa käytetään tarvittaessa helpottamaan ja tukemaan toimien osa-alueita.

3.2.3. Hankkeelta odotetut tulokset

Hankkeen odotetaan luovan valmiuksienkehittämisen- ja osallistamistoimia koskevia malleja, joita kouluttajat ja akateemisten opetussuunnitelmien kehittäjät voisivat toisintaa, jotta tulevat tekoälyalan toimijat saataisiin tietoisiksi ongelmista, jotka liittyvät siihen, että vastuuttomat toimijat käyttävät siviilitekoälyä tarkoituksenvastaisesti ja väärinkäyttävät sitä, ja miten he voivat auttaa lieventämään näitä ongelmia vastuullisen innovoinnin prosessien avulla.

Kun hanketoimet on saatu päätökseen, osallistujien (kouluttajien mutta myös STEM-aineiden opiskelijoiden) odotetaan pystyvän käyttämään ja edistämään vastuullisen innovoinnin välineitä, menetelmiä ja konsepteja, joilla tunnistetaan ja lievennetään tekoälyalan siviilitutkimuksen ja -innovoinnin kehittämiseen ja levittämiseen liittyviä riskejä, joita aiheutuu tarkoituksenvastaisesta käytöstä ja väärinkäytöstä.

Hankkeen toimien odotetaan myös luovan verkoston kouluttajista, opetussuunnitelmien kehittäjistä ja opiskelijoista, jotka paitsi edistäisivät hanketoimia tekoälykoulutuksessa ja ammatillisissa yhteisöissä (esim. kansainvälisen sähkö- ja elektroniikkainsinöörien järjestön tietoteknistä älykkyyttä käsittelevän yhteisön konferensseissa) myös edistäisivät valtiojohtoisten kansainvälisten hallintoprosessien teknisiä valmiuksia (esim. eräitä tavanomaisia aseita koskeva yleissopimus, joka koskee uusia teknologioita tappavien autonomisten asejärjestelmien alalla).

Näiden toimien lyhyen ja pitkän aikavälin arvo osoitetaan toimintaa edeltävillä ja sen jälkeisillä kyselytutkimuksilla.

3.3. Hanke 3 – Vastuullinen tekoälyinnovointi rauhaa ja turvallisuutta varten sekä sen pidempiaikaisen kestävä kehityksen edistäminen, levittäminen ja vaikutukset

3.3.1. Hankkeen tarkoitus

Hankkeen nro 3 tarkoituksena on helpottaa vastuullisen tekoälyinnovoinnin pidempiaikaisen kestävä kehityksen edistämistä, levittämistä ja vaikutuksia keinona lieventää rauhaan ja turvallisuuteen liittyviä riskejä, joita tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö ja väärinkäyttö voivat aiheuttaa. Tähän pyritään järjestämällä keskusteluja tekoälyteollisuuden kanssa, käymällä useiden sidosryhmien välistä vuoropuhelua, laatimalla julkinen raportti ja toteuttamalla kohdennettua tiedottamista. Hankkeella pyritään varmistamaan, että tuotettu työ, erityisesti koulutus-, valmiuksienkehittämis- ja osallistamistoimet, tavoittaa tekoälyalan yhteisön ja vaikuttaa siihen yleisesti kaikilla tasoilla (opiskelijoista insinööreihin ja muihin tekoälyn ammattilaisiin) ja maantieteellisten, alakohtaisten ja muiden rajojen yli. Laajan ja syvän vaikutuksen mahdollisuuksien lisäämiseksi on olennaisen tärkeää tehdä yhteistyötä kyseessä olevalla alalla toimivien ammattialajärjestöjen, kuten kansainvälisen sähkö- ja elektroniikkainsinöörien järjestön, kanssa ja tehdä moniulotteista osallistamista tiedeyhteisön, teollisuuden ja muiden alojen kesken.

Tällaiset toimet antavat tekoälyalan eri yhteisöjen asiasta kiinnostuneille edustajille mahdollisuuden ottaa ongelmaa koskevaa vastuuta ja esittää omia näkemyksiään siitä, miten riskien lieventämistoimia voidaan toteuttaa ja edistää kestäväällä tavalla maailmanlaajuisessa tekoälyalan yhteisössä ja sen eri osien välillä. Hankkeen pitkän aikavälin arvon kannalta valtioille, hallitustenvälisille järjestöille ja muille tahoille on myös tärkeää, että tekoälyalan toimijat voivat oppia valtiollisilta asiantuntijoilta, jotka osallistuvat riskien lieventämiseen aseriisunta-, asevalvonta- ja asesulkualalla, ja tehdä yhteistyötä heidän kanssaan. Kestävyyden kannalta on myös ratkaisevan tärkeää varmistaa, että osallistamistoimien tuottamat näkemykset analysoidaan, konsolidoidaan ja levitetään asianmukaisesti.

3.3.2. Hankkeen kuvaus

Hanke koostuu seuraavista keskeisistä osa-alueista:

- a) usean sidosryhmän välinen vuoropuhelu aiheesta ”Vastuullinen tekoälyinnovointi rauhaa ja turvallisuutta varten” (korkeintaan 9 kokousta) – Tässä virtuaalisena vuoropuheluna järjestettävien kokousten sarjassa kokoontuvat tiedemaailman, tutkimuksen, yksityisen sektorin ja perinteisen asevalvonnan asiantuntijat EU:sta ja muualta maailmasta keskustelemaan seuraavista:

- i. teknologiset suuntaukset, jotka voivat aiheuttaa kansainväliseen rauhaan ja turvallisuuteen vaikuttavia riskejä, jotka koskevat levittämistä, tarkoituksenvastaista käyttöä ja väärinkäyttöä;
- ii. keinot, joilla osallistua riskien lieventämiseen vastuullisten innovointiprosessien, -menetelmien ja -välineiden avulla sekä sidosryhmäyhteisöjen, mukaan lukien muilla aloilla, kuten bio- ja kemianteollisuudessa toimivat yhteisöt, väliseen vuoropuheluun ja tietämyksen jakamiseen liittyvät mahdollisuudet ja haasteet; ja
- iii. itsenäisen asiantuntijaverkoston ja vuoropuheluun liittyvien toimien mahdollinen arvo, tarkoitus ja muoto. Asiantuntijaryhmä kokoontuu useita kertoja vuodessa ja pyrkii järjestämään kaksi julkista laajemmalle yhteisölle suunnattua tapahtumaa.

Yhdeksästä virtuaalikokouksesta kahden on tarkoitus olla avoimia yleisölle laajemman kuulemisen helpottamiseksi.

- b) yksityisen sektorin asiantuntijakeskustelut (enintään 6) – Tässä virtuaalisesti käytävien asiantuntijakeskustelujen sarjassa käynnistetään vuoropuhelu vastuullista tekoälyä koskevien innovointiprosessien parissa työskentelevien yksityisen sektorin toimijoiden kanssa (esim. tekoälyä koskeva kumppanuus). Käsiteltävänä on se, miten ne voivat osaltaan minimoida riskejä, joita aiheutuu siitä, että vastuuttomat toimijat käyttävät tekoälyalan siviiliteknologiaa tarkoituksenvastaisesti ja väärinkäyttävät sitä, sekä tutkia mahdollisia kannustimia yksityisen sektorin kehittämiseen tätä varten. Aiheina ovat muun muassa
 - i. kansainvälisen turvallisuuden ja aseriisunnan merkitys yksityissektorille;
 - ii. oikeudellinen ympäristö (tai ympäristöt), jo(i)ssa tekoälyä kehitetään, otetaan käyttöön ja käytetään eri puolilla maailmaa;
 - iii. se, miten voidaan hyödyntää riskinarviointimekanismeja ja muita toimenpiteitä, jotka ovat osa vastuullisia innovointiprosesseja ja yritysten vaatimustenmukaisuusohjelmia tai jotka voitaisiin sisällyttää niihin; ja
 - iv. muista asevalvontaan liittyvistä teollisuudenaloista, prosesseista ja kehyksistä saadut kokemukset (esim. bio- ja kemianteollisuus).
- c) tekoälyalan yhteisölle ja asevalvontayhteisölle suunnattu raportti tekoälyalan yhteisön näkökulmista asevalvontaan ja riskien lieventämiseen tekoälyalalla (1) – Tämän raportin laadinnassa kootaan ja konsolidoidaan hankkeen havainnot ja suositukset yhdeksi viiteasiakirjaksi, joka on tarkoitettu sekä tekoälyalan siviiliyhteisölle että asevalvontayhteisöille. Raportissa käsiteltäisiin sitä, miten voidaan tunnistaa, arvioida ja käsitellä kansainväliseen rauhaan ja turvallisuuteen liittyviä riskejä, joita tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö ja väärinkäyttö voivat aiheuttaa.
- d) tiedonlevitystapahtumat, joiden tarkoituksena on kuulla tekoälyalan yhteisöä ja asevalvontayhteisöä ja yhteistyö niiden kanssa (päätetään myöhemmin) – Ryhmä etsii mahdollisuuksia tiedottaa työstä ja sitä koskevista havainnoista ja suosituksista koko hankkeen ajan. Tapahtumien muoto ja esitysten sisältö räätälöidään kohderyhmien tarpeiden mukaan. Näihin voivat kuulua asesulku- ja asevientityöryhmän kokoukset, eurooppalainen tekoälyallianssi; tappavien autonomisten asejärjestelmien alan uusia teknologioita käsittelevä hallitustenvälinen asiantuntijaryhmä; tekoälyä käsittelevä virastojen välinen työryhmä

(IAWG-AI); Kansainvälisen televiestintäliiton (ITU) ”AI for Good”-aloite; YK:n aseidenriisunnan tutkimuslaitoksen (UNIDIR) vuotuinen innovointivuoropuhelu; ja kansainvälinen sähkö- ja elektroniikkainsinöörien järjestö. Ryhmä pyrki myös kahdenväliseen yhteistyöhön hallitusten, tiedemaailman ja yksityisen sektorin asiaankuuluvien sidosryhmien kanssa.

3.3.3. Hankkeelta odotetut tulokset

Hankkeen odotetaan luovan neuvoston päätöksen välittömän keston jälkeenkin vaikuttavan perustan sellaisten tekoälyprosessien vastuullisen innovoinnin kestävälle kehitykselle, levittämiselle ja vaikutusmahdollisuuksille, joilla puututaan leviämiseen, tarkoituksenvastaiseen käyttöön ja väärinkäyttöön liittyviin riskeihin ja niiden vaikutuksiin rauhaan ja turvallisuuteen.

Sidosryhmien välisen vuoropuhelun odotetaan tarjoavan mallin riskien lieventämistä koskevalle tiedonvaihdolle ja yhteistyölle paitsi maailmanlaajuisessa tekoälyalan yhteisössä myös tekoälyalan siviiliyhteisön ja aseriisunta-, asevalvonta- ja asesulkuyhteisöjen välillä. Tällaista mallia voitaisiin käyttää poliittisten päättäjien tutustuttamisessa tekoälyn vastuullisen innovoinnin kannalta merkityksellisiin ja keskeisiin teknologisiin ja tieteellisiin edistysaskeliin ja lisäksi teknisen alan yleisön tutustuttamiseksi poliittisten päättäjien tämänhetkiseen toimintaympäristöön. Hankkeen odotetaan edistävän kestäviä suhteita ja toimintaa kiinnostuneiden toimijoiden välillä näissä eri yhteisöissä ja niiden välillä. Tällaisten heterogeenisten verkostovaikutusten odotetaan mahdollistavan rauhaa ja turvallisuutta edistävän vastuullisen tekoälyinnovoinnin kattavamman kehittämisen ja laajemman edistämisen myös hankkeen keston jälkeen.

Yksityisen sektorin vuoropuhelun odotetaan mahdollistavan tekoälyalan yksityisen sektorin laajemman ja syvemmän osallistumisen sellaisten rauhaa ja turvallisuutta koskevien riskien tunnistamiseen, ehkäisemiseen ja lieventämiseen, joita tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa.

Hankkeen odotetaan lisäävän yksityisen sektorin prosessien keskeisten toimijoiden ymmärrystä ja vastuunottoa liittyen ongelmiin, joihin hankkeella pyritään puuttumaan. Lisäksi sillä pyritään helpottamaan vastuullisten innovointiprosessien, -menetelmien ja -keinojen (laajempaa) käyttöönottoa ja toteuttamista nykyisissä organisaation riskinhallintamekanismeissa ja -menettelyissä.

Sidosryhmien välisen vuoropuhelun ja yksityisen sektorin asiantuntijakeskustelun odotetaan myös tuottavan näkemyksiä useista merkittävistä kysymyksistä, kuten a) siitä, miten vastuullisia innovointimenetelmiä ja -keinoja voidaan edelleen kehittää ja ottaa käyttöön, jotta voidaan tunnistaa, ehkäistä ja lieventää riskejä, joita tekoälyalan siviilitutkimuksen ja -innovoinnin tarkoituksenvastainen käyttö tai väärinkäyttö voivat aiheuttaa; b) miten tekoälyn tutkimusta ja innovointia voidaan käyttää myönteisesti rauhan ja turvallisuuden tavoitteiden tukemiseen (esim. konfliktien varhaisvaroitusta ja humanitaarista apua koskevat soveltamistavat); ja c) miten voidaan edistää kattavampaa vuoropuhelua ja tiedonvaihtoa tekoälyalan siviiliyhteisön toteuttamien riskinvähentämistoimien (eli erilaisten IEEE-johtoisten aloitteiden) ja hallitustenvälisellä tasolla aseriisunta-, asevalvonta- ja asesulkuyhteisöissä jo käynnissä olevien toimien välillä.

Raportoinnin ja tiedonlevitystoimien avulla analysoidaan, vakiinnutetaan ja levitetään hankkeiden nro 1, nro 2 ja nro 3 kautta saatuja näkemyksiä ja tuetaan näin hanketoimien tulosten ja suositusten tunnetuksi tekemistä maailmanlaajuisessa tekoälyalan yhteisössä ja poliittisessa yhteisössä. Niiden odotetaan myös auttavan varmistamaan vaikutusten kestävyys vielä hankkeiden keston jälkeen.

4. KESTO

Hankkeiden toteutuksen arvioitu kokonaiskesto on 36 kuukautta.