Supporting the design of human rights compliant models for the use of artificial intelligence for counter-terrorism.

The project will strengthen Member States, technology companies, and practitioners' understanding of how AI can be used in a human rights compliant manner for counter-terrorism purposes and provide actionable input on human rights considerations in national and regional development of policies and regulations on the use of AI in the context of counter-terrorism.

Budget
$2,236,536.00

Pillar
Pillar IV -Ensuring Human Rights and the Rule of Law

Project Status
Concept

Implementation Period
2021-2023

UN Org
OHCHR

UN Global CT Compact Partners
UNICRI, UNOCT

Countries
Global

Current Donors
This project is currently unfunded.

Description of the Project

The project seeks to strengthen practical understanding of how AI can be used in a human rights compliant manner for counter-terrorism purposes and to provide practical input on human rights considerations to ongoing or incipient national or regional discussions on policies and regulations on the use of AI in the security context. Ultimately, it is expected that the project will contribute to ensuring that Member States use AI technologies to counter-terrorism in compliance with their human rights obligations. Building on the OHCHR, UNICRI and UNOCT/UNCCT joint research project on Human Rights Aspects to the Use of Artificial Intelligence for Counter-Terrorism, this project seeks to strengthen consideration of the human rights compliant AI for counter-terrorism by engaging with both the private sector entities developing these technologies and the regulatory bodies, which are creating the polices and regulations that frame both their development and eventual use in an effort to operationalize the principles/standards identified by in the research.

Through the project, **the private sector will be engaged in order to stress test AI applications developed for Member States for use in counter-terrorism operations and guidance and support will be provided to address any potential inadequacies in terms of human rights that were identified**. Moreover, regulatory bodies actively exploring policies or regulations on AI will be engaged in order to ensure that human rights principles and standards for the use of AI in counter-terrorism are reflected. Awareness-raising and outreach activities targeting sets of stakeholders, private sector and regulatory bodies, will be organized to advance the project objectives and activities.

Rationale

Recent technological advancements have sparked a growing interest from Member States, the counter-terrorism community and technology companies to explore how AI can be used in counter-terrorism efforts. The exponential growth of data concerning terrorist groups and individuals, boosted by the widespread use of the internet and social media and the overall increased digitalization of societies around the world, has heightened such interest. AI has the potential to analyze and learn from vast amounts of data and identify correlations previously unknown to humans. Numerous public and private sector initiatives are thus researching and developing the use of AI-enabled tools to, inter alia, moderate content; gather and share information about individuals; identify individuals vulnerable to radicalization; map and analyze social networks; predict threats and attacks; and enhance surveillance capabilities, including through facial recognition technologies. Although tools such as these can be a powerful instrument to support counter-terrorism and law enforcement agencies, they can also impact human rights significantly.

The UN Security Council and General Assembly have both recognized the value and utility of such kinds of technology, while also underscoring that it needs to be deployed in compliance with international law, in particular with international human rights law. For example, Security Council Resolution 2396 (2017) stressed the need for cooperation with the private sector and civil society in the development and implementation of more effective means to counter terrorism through innovative technological solutions "all while respecting human rights and fundamental freedoms in compliance with domestic and international law". General Assembly Resolution 73/174 on Terrorism and Human Rights expressed concern about terrorist use of technology and urged States to take "preventative measures" (which could include the use of AI technology) "while acting in full compliance with their obligations under international law." Furthermore, in his 2018 report to the General Assembly, the UN Special Rapporteur on Freedom of Expression further examined the impact of artificial intelligence (AI) on the information environment and provided recommendations that Member States and companies make human rights central to the "design, deployment, and implementation of artificial intelligence systems." Finally, the project is aligned with the growing digital agenda of the United Nations as well as related human rights concerns, as reflected by the Secretary-General's Call to Action for Human Rights and recommendations of the High-level Panel on Digital Cooperation.

Many of these human rights impacts have been identified in the ongoing OHCHR, UNICRI and UNOCT/UNCCT joint research project on Human Rights Aspects to the Use of Artificial Intelligence for Counter-Terrorism, including AI's potential to interfere with privacy and data protection standards and exacerbate discrimination due to biases present in data or restrict the freedom of expression. To properly harness the potential of AI in counter terrorism, it is paramount that Member States ensure that the use of this technology is responsible in that it reflects and respects human rights and fundamental freedoms. If they fail to do so, the evidence generated may be deemed inadmissible in courts and the public trust in the counter-terrorism and law enforcement agencies will be eroded.

Outcomes

This project seeks to achieve the following outcomes:

Outcome 1: Increased United Nations engagement with the private sector on operationalizing human rights in AI for counter-terrorism purposes.
Outcome 2: A human rights focused stress test of AI applications developed by private sector entities for the counter-terrorism purposes.
Outcome 3: Mapping of AI standard setting initiatives around world.
Outcome 4: Application of international human rights law principles and standards by regulatory bodies reviewing the use of AI in counter-terrorism contexts.
Outcome 5: Increased acceptance and understanding of the specific human rights standards applicable to the use of AI in counter-terrorism.
Outcome 6: Engagement with regulatory bodies on the inclusion of principles/standards identified in the OHCHR, UNOCT and UNICRI report 'Human Rights Aspects to the Use of Artificial Intelligence for Counter-Terrorism.'
Outcome 7: Capacity-building guidance and support for private sector entities and regulatory bodies on addressing specific human rights concerns in the use of AI in counter-terrorism.

Outputs/Indicative Activities

The project will implement the following activities:

- Review and update findings from OHCHR, UNOCT and UNICRI project on human rights as appropriate.
- Develop a roadmap for operationalizing the principles/standards set out in the report.
- Organize outreach and awareness-raising activities targeting the private sector.
- Develop a questionnaire based on the principles/standards specifically for use by private sector entities.
- Liaise with private sector to secure engagement in utilising the assessment questionnaire as part of a stress test of their AI applications.
- Provide guidance and support to private sector entities to address any potential human rights concerns identified and on operationalizing the principles/standards.
- Identify and map AI standard-setting initiatives around the world.
- Liaise with regulatory bodies considering AI policies or regulation and advocate for inclusion of principles/standards set out in the report in these initiatives.
- Provide guidance and support to regulatory bodies on the basis of the principles/standards and in order to address specific human rights concerns identified through engagement with the private sector.

Human Rights Mainstreaming

Ensuring human rights compliance in the use of AI in counter-terrorism is the very core and desired outcome of the project and, in this regard, plays a key role throughout the implementation. The project will seek to build upon the work conducted by OHCHR/UNICRI/UNOCT through the project Human Rights Aspects to the Use of Artificial Intelligence for Counter-Terrorism, as well as the Expert Seminar on Artificial Intelligence and the Right to Privacy organized by OHCHR in May 2020, where it addressed the recommendation of the experts that oversight mechanisms need "to be adequately resourced to monitor the effects of AI technologies on human rights, and have access to AI experts to support their mandates." It also directly takes up the findings of the UN Special Rapporteur on Freedom of Expression in his 2018 report to the General Assembly on the impact of artificial intelligence (AI) on the information environment, in which recommendations were provided that States and companies make human rights central to the "design, deployment, and implementation of artificial intelligence systems." It will furthermore take into consideration the recommendations issued by the Committee on the Elimination of Racial Discrimination in its guidance to combat racial profiling, which emphasized, among other issues, the serious risk of algorithmic bias when AI is used in law enforcement.

Gender Mainstreaming

Regarding gender mainstreaming, the project team will assess the gendered dimensions of the use of AI in counter-terrorism, highlighting the potential differentiated effects of the use of AI in counter-terrorism on men and women, including adverse effects on women and identity solutions. In terms of participation, a balanced ratio of male to females will be encouraged and all-male panels in conferences and workshops will not take place within the project. Gender-based indicators will be developed and included in the logical framework and gender-related data will be continuously monitored throughout implementation, supporting a gender analysis as part of the project's overall evaluation.