



STRATEGIC PLAN 2020-2025

STRATEGIC PLAN

2020-2025

UNICRI CENTRE FOR AI AND ROBOTICS

INTRODUCTION

In response to rapid advancements in digital technologies and the surge in data being produced by both humans and machines, the United Nations Interregional Crime and Justice Research Institute (UNICRI) established a specialized Centre for Artificial Intelligence and Robotics (hereinafter ‘the Centre’) in September 2017.

Located in the Hague, the Netherlands, the Centre was established with the support of the Municipality of the Hague, the Ministry of Foreign Affairs of the Kingdom of the Netherlands and strategic partners from the private sector.

This five-year strategic plan, covering the period of 2020 to 2025, is intended to provide a framework for the Centre and its operations during this period, which will serve to inform the focus of its programming and guide how it interacts with donors and stakeholders.

MISSION:

The Centre advances understanding of artificial intelligence (AI) and unmanned robotic vehicles with a special focus on the aspects of crime, justice and other threats to security in order to promote public safety and reduce crime¹.

Through research, multi-stakeholder discussions and awareness-raising, the Centre supports national authorities and relevant representatives within the criminal justice system (law enforcement, courts and corrections), to understand both the opportunities and potential pitfalls presented by the use of these technologies, as well as the threat and risk of their malicious use.

¹ In order to ensure a holistic approach, interlinked emerging technologies that are of pertinence to the application of AI and unmanned robotic vehicles, including the Internet of Things, biometrics, environmental sensors, blockchain and quantum technologies, also fall within the remit of the Centre’s mission.

The Centre furthermore sees it as part of its mission to advance understanding of how the responsible use of these technologies can play a role in the realisation of the goals set out in the [2030 Agenda on Sustainable Development](#), in particular [Goal 16](#) on promoting peaceful and inclusive societies, providing access to justice for all and building effective, accountable and inclusive institutions at all levels. The Centre will furthermore seek to contribute to relevant elements of the Secretary-General's Roadmap for Digital Cooperation.

This mission of the Centre is in line with UNICRI's overarching [Strategic Programme Framework](#), specifically the fourth strategic priority "*Security through Research, Technology and Innovation*", in that it contributes to advancing knowledge and understanding of the global impacts, opportunities and challenges of advancements in AI, in particular from a criminal justice perspective, and provides support to Member States, international organizations, the private sector, civil society, the scientific and academic community, and other relevant stakeholders on keeping pace with technological change.

VISION:

The Centre envisions a world where crimes are prevented, investigated and prosecuted and offenders are deterred and rehabilitated through the criminal justice system more effectively and efficiently through the use of AI, robotics and related new technologies in a manner that does not erode human rights, deepen inequality or exacerbate discrimination.

With a strong emphasis on the lawful and human rights compliant use of these technologies, the public trusts both the criminal justice system and the technology that it applies to safeguard them. At the same time, authorities are aware of and are better prepared to prevent and respond to the possible malicious use of these technologies for criminal or terrorist purposes.

In this world, the Centre is a hub or platform leading national, regional and international discussions on these technologies from the perspective of UNICRI's mandate. It is perceived by its partners and stakeholders as a trustworthy source of knowledge and its work is recognized for its substantive value, as well as neutral and rights-based stance.

PRIORITIES AREAS AND PERSPECTIVES:

New and emerging technology is a broad and expanding thematic, replete with a multitude of different dimensions and angles. Precisely how these technologies interact with matters related to crime, terrorism and other threats to security is however an underexplored domain.

In order to provide guidance for the Centre's programming for the coming five year period, the several priority areas have been identified at both governance and operational levels. The priorities, place the Centre at the forefront of:

- Building knowledge on the possible **malicious use** of AI by criminals and terrorist groups, as well as potential counter-measures

- Enhancing awareness of the threats programmatically generated or manipulated voice or video content, such as **deepfakes**
- Fostering **responsible AI innovation** within the law enforcement community
- Promoting and supporting the development of policy frameworks for the deployment of **facial recognition software and other AI surveillance technologies**
- Exploring the development of pilot AI applications in criminal investigations, in particular to combat the rise in **online child sexual abuse materials**
- Enhancing **cybersecurity** through the use of AI to support the detection and investigation of and protection from cyberattacks
- Building knowledge on the use of **AI in counter-terrorism**, in particular for the identification of terrorist use of social media and the internet
- Advancing discussion on **drone** use and policy, as well as **counter-drone measures**
- Analysing the possible application of AI in the **administration of criminal justice and corrections administration**

This list of priorities is not exhaustive and may be refined or expanded in consultation with the AI Technical Advisory Group described below, catering for technical and political developments globally, as well as the shifting needs and requirements of the Centre. With respect to geographical scope for activities implemented in accordance with these priority areas and perspectives, the Centre will endeavour to widely engage and strongly support developing countries.

In addition to these priority areas, the work of the Centre will be tailored to address certain cross-cutting perspectives have also been identified as meriting a high degree of priority, namely:

- Human rights
- Ethics
- Gender
- Digital divide
- Culture
- Public trust
- Social acceptance
- Technology

GUIDING PRINCIPLES:

Advancements in emerging technologies, in particular AI, have generated considerable discussion on the principles, values or virtues that should guide the design and use of these technologies. Building upon this, the Centre has identified the following principles and

associated requirements as fundamental for its approach to the use of AI, robotics and related technologies.

At the highest level, respect for internationally recognized body of human rights law, the rule of law, and the principle of democracy are considered as the overarching framework for the design and use of these technologies. In this regard, any system developed must conform with the maxim of *primum non nocere* – do no harm.

To achieve these principles, and to justify the placing of public trust in the use of these technologies within the criminal justice system, the following requirements must be ensured throughout the entire lifecycle of the system:

- Fairness
- Accountability
- Transparency
- Explainability

At all times, such systems must also be human-centric, seeking to augment and not replace humans, by pairing human judgment and empathy with the data-processing capabilities of machines.

In light of the fact that data is the fuel that drives any and all AI systems, the Personal Data Protection and Privacy Principles of the United Nations complement these principles and requirements and serve as a basic framework for the processing of all personal data. These include the principles of fair and legitimate processing; purpose specification; proportionality and necessity; retention; accuracy; confidentiality; security; transparency; transfers; and, accountability².

Collectively, these serve as the overarching guiding principles for the Centre, with respect to the design, development and deployment of AI, robotics, and related technologies and are interwoven into all its endeavours.

EXECUTION OF THE PLAN:

The Centre will endeavour to develop and implement quick impact projects – small-scale, cost-effective projects that are planned and implemented within a short timeframe – to address the aforementioned priorities and perspectives, building knowledge to serve a baseline for possible future actions by other stakeholders.

These projects will employ several approaches and tools, including:

- Action-oriented research

² <https://www.unsystem.org/personal-data-protection-and-privacy-principles>

- Knowledge development/dissemination
- Training and technical workshops
- Advocacy with policy- and decision-makers
- Proof of concept development and piloting of tools

Network building and the creation of strategic partnerships has also been identified as being a fundamental part to the modus operandi of the Centre's activities and is integral to its success. In this regard the Centre has built an extensive international network of partners that it engages for its activities and to convene expert-level meetings, training courses and workshops worldwide, as well as high-level visibility events.

This includes, Europol, INTERPOL, the Office of the High Commissioner for Human Rights (OHCHR), Organization for Security and Co-operation in Europe (OSCE), the United Nations Office of Counter Terrorism (UNOCT), the United Nations Office of Drugs and Crime (UNODC), the United Nations Office of Information and Communications Technology (OICT), the World Economic Forum – Centre for the Fourth Industrial Revolution, as well as law enforcement agencies, academic institutions and industry partners throughout UN Member States.

With respect to its UN partners, the Centre will endeavour to make itself a resource and source of expertise that is available to the entities within UN system, as well as to rely existing expertise throughout the system.

AI TECHNICAL ADVISORY GROUP:

Given the highly technical nature of the thematic and the rapid pace of technological innovation, keeping abreast of the latest developments will be a significant ongoing challenge for the Centre in the pursuit of its mission.

In order to support the Centre to do so and ensure that its programming is relevant, sensitive to and reflects the latest technological advancements, as well as related political and societal developments, an *AI Technical Advisory Group* (AITAG) will be established.

This informal and independent group, will consist of recognized technical experts from industry and academia, relevant stakeholders that may be considered as the end-users in national authorities in Member States, as well as relevant experts from within the United Nations system and other international and civil society organizations that form the community of concerned stakeholders. Due regard will be paid to gender balance and to the principle of equitable geographical distribution in the selection of its members.

The group will meet on an ad hoc basis, performing a purely advisory function. Nothing provided by the group will be considered binding on the Centre.

FINANCING THE CENTRE

The Centre is funded on the basis of voluntary contributions by donors.

Recognizing this, fundraising and donor relations in general is of paramount importance for realization of the present strategic plan and to ensure sustainability of results achieved through its ongoing work.

In this regard, the Centre has built several strategic partnerships with donors, including most notably, the Kay Family Foundation and 1QB Information Technologies, that have contributed to sustaining the work of the Centre. During the forthcoming five-year period, the Centre will continue to actively work to further identify potential donors and jointly develop and implement projects in line with the above-described priority areas.

In connection with fund-raising, the Centre will actively engage in visibility and outreach activities, targeting different audiences, this includes a dedicated page on the UNICRI website, a strong social media presence, factsheets and e-flyers, public speaking engagements, interventions at conferences, organisation of various events, and participation in meetings organized by other stakeholders, as well as publications and reports, which also overlap as outputs of the Centre's work.

In all visibility and outreach activities, the support and contributions of existing donors and strategic partners will be recognized.

