



UNITED ARAB EMIRATES  
MINISTRY OF INTERIOR

## The AI for Safer Children Initiative Ethical and Legal Report

V1.0 – December 2021

### EXECUTIVE SUMMARY

The *AI for Safer Children* initiative, implemented in partnership between the United Nations Interregional Crime and Justice Research Institute (UNICRI), through its Centre for Artificial Intelligence and Robotics, and the Ministry of Interior of the United Arab Emirates (UAE), aims to support global law enforcement in exploring the positive potential of artificial intelligence (AI) to tackle child sexual exploitation and abuse online. To that end, the initiative envisages:

- Building a web-based platform – the “AI for Safer Children Global Hub” – where law enforcement agents can access a database of AI tools that are available to combat online child sexual exploitation and abuse, as well as communicate with each other and learn how to use these tools responsibly.
- Enhancing knowledge on the issue of online child sexual exploitation and abuse and the role that AI can play in supporting law enforcement in combating it, through awareness-raising and outreach activities.

As an initiative co-developed by UNICRI, and United Nations (UN) entity, the *AI for Safer Children* strives to uphold the values and principles enshrined in the UN Charter and in the Universal Declaration of Human Rights. This is particularly the case given that the initiative engages with what can be considered highly sensitive subjects, namely the use of AI in law enforcement and child sexual exploitation and abuse. UNICRI and the Ministry of Interior of the UAE have thus established an ethical and legal process of the *AI for Safer Children* initiative to ensure that the initiative is ethically and legally developed from the outset.

Under this process, the decision was taken to prepare an Ethical and Legal Report to ultimately guide UNICRI and the Ministry of Interior of the UAE in the implementation of the initiative. This document corresponds to the first version (V1.0) of this report, prepared through extensive desk-based research and numerous dedicated interviews and meetings with ethics experts, law enforcement agents, representatives of technology providers and representatives from other international organizations. This version is intended for discussion with a group of independent ethics experts in a virtual Ethics Expert Meeting taking place on 7 December 2021.

This first version of the *AI for Safer Children* Ethical and Legal Report outlines four core principles guiding the *AI for Safer Children* initiative.

#### Core principles guiding the *AI for Safer Children* initiative [\[read more\]](#)

1. The values enshrined in both the UN Charter and the Universal Declaration of Human Rights and human rights principles at large are at the forefront of the *AI for Safer Children* initiative.
2. The *AI for Safer Children* initiative will foster the responsible development and use of AI in law enforcement, in line with UN values.

3. The *AI for Safer Children* initiative promotes global inclusion targeting all UN Member States and gathers perspectives from a diverse range of stakeholders, including those at risk of being left furthest behind.
4. The values of integrity, impartiality, independence and transparency will be upheld in all aspects of the development of the *AI for Safer Children* initiative.

This report then explains the ensuing concerns regarding the initiative's activities, and the concrete actions that either are being taken or are planned to be taken to practically instil those principles and address the ethical and legal concerns.

#### **The ethical and legal considerations and concrete measures to address them:**

##### **Promoting global inclusiveness in a digitally divided world** [\[read more\]](#)

- Raising awareness about the issue of online child sexual exploitation and abuse and its global nature along with the limitations caused by the digital divide, notably the reduced amounts of data available for certain countries and regions, particularly in the Global South.
- Ensuring geographical diversity of the participants and speakers at the events.
- Promoting the initiative globally and taking into account the perspectives and needs of law enforcement agents from countries across all regions in the development of the initiative, particularly from those at risk of being left behind.
- In the design of the Global Hub, writing the texts in accessible language, e.g. avoiding over-complicated phrasing or unnecessarily technical language, and translating subsequent versions of the platform from English into at least other official UN languages.
- Including a learning section on the Global Hub to build the capacity of law enforcement agents worldwide with different levels of technical proficiency to responsibly use AI in their work.
- Researching AI tools developed in different countries across all regions to be featured on the Global Hub.
- Adding, in the future, a national and regional training component to the initiative.
- Collecting information from the technology providers of the AI tools to be featured on the Global Hub about which languages the tool is available in and, for those with text analysis capabilities, which languages the tool can analyze.

##### **Gender mainstreaming** [\[read more\]](#)

- Raising awareness about the risks of gender biases in the development of the AI tools (for instance, because of under-representation of women and girls in the training data), with the consequent negative impact on women and girls if used by law enforcement without regard to this bias. This would be achieved by including the information on the Global Hub's learning section as well as highlighting it in advocacy and outreach events.
- Encouraging law enforcement agencies to ensure gender balance of the law enforcement agents they designate as their representatives on the Global Hub.
- Ensuring gender diversity of the participants and speakers at meetings and events.
- Ensuring gender diversity of the team developing the AI for Safer Children initiative.

##### **Protection of personal data processed within the initiative** [\[read more\]](#)

- Implementing and following a transparent data processing policy in line with the UN Personal Data Protection and Privacy Principles.
- Utilizing an official, secure server on which to store collected data.

##### **Security of the AI for Safer Children Global Hub** [\[read more\]](#)

- Hosting the Global Hub under the supervision of the United Nations Office of Information and Communications Technology (OICT), which has experience with other secure websites and the gambit of possible cybersecurity mechanisms.

- Limiting access to the Global Hub to law enforcement agents through, for instance, a secure log-in and possibly requiring further verification of their identity and affiliation.

#### **Avoid featuring unethical or unlawful tools** [\[read more\]](#)

- Being clear and transparent about the core principles and scope of the AI for Safer Children initiative, both by providing a comprehensive statement as such on the Global Hub and by making publicly available documents such as the present Ethical and Legal Report.
- Collecting information from the technology providers on the ethical considerations involved in their development of the AI tools featured on the Global Hub and providing this information to the law enforcement users.
- Having processes in place to ensure that, if reliable information that any tool raises ethical and legal issues is received, such tools are removed from the Global Hub and its users are informed.

#### **Promote the responsible use of AI by Law Enforcement** [\[read more\]](#)

- Providing the information collected from the technology providers about the ethics behind their AI tools to the law enforcement users.
- Being clear and transparent about the core principles and scope of the AI for Safer Children initiative.
- Including guidance on the Global Hub's learning section specifically dedicated to the responsible use of AI, from developing a tool in-house to the considerations of its integration into the workforce and deployment.

#### **Fair promotion of AI providers** [\[read more\]](#)

- Ensuring that the Global Hub does not favour any AI tool, for instance by giving every tool equal visibility on the platform and not allowing any kind of commercial advertisement.
- Maintaining an impartial relationship with technology providers by not favouring any specific provider as a speaker or participant at events and activities related to the initiative.
- Proactively researching the AI tools to be featured on the Global Hub instead of relying on private industry to reach out to the initiative, thereby preventing bigger technology providers from being overrepresented.

The relevant feedback from the experts during the Ethics Expert Meeting will be considered in the preparation of a second version, which will guide UNICRI and the Ministry of Interior of the UAE in the implementation of the initiative. If needed, subsequent versions of the report will be prepared and discussed in follow-up ethics expert meetings.

The considerations and proposed measures outlined above, as well as their exact configuration, are thus not set in stone and will be expanded and improved with the external views collected during and following the Ethics Expert Meeting on 7 December 2021. At the same time, the process of identifying and addressing the ethical and legal considerations of the *AI for Safer Children* is an ongoing process that will be ever evolving throughout the initiative. This means that the ethical and legal report is not a one-off deliverable. On the contrary, any lessons learned, supervening relevant guidelines or regulatory frameworks or related work from UNICRI will feed into this process, to ultimately contribute to the best possible outcome: supporting law enforcement globally to responsibly use AI in combatting online child sexual exploitation and abuse.

## Introduction

Implemented in partnership between the United Nations Interregional Crime and Justice Research Institute (UNICRI), through its Centre for Artificial Intelligence (AI) and Robotics, and the Ministry of Interior of the United Arab Emirates (UAE), the *AI for Safer Children* initiative aims to support law enforcement and related authorities in exploring the positive potential of AI to tackle child sexual exploitation and abuse online.<sup>1</sup>

To that end, the initiative envisages building and piloting a web-based platform – the “AI for Safer Children Global Hub” – where law enforcement agents can access a database of AI tools that are available to combat online child sexual exploitation and abuse, as well as communicate with each other and learn how to use these tools responsibly. In addition to this, the *AI for Safer Children* comprises awareness-raising and outreach activities on the issue of online child sexual exploitation and abuse, including organizing and participating in events for visibility, network-building and advocacy on this subject.

As an initiative under the auspices of the United Nations (UN), the *AI for Safer Children* needs to uphold the values and principles enshrined in the UN Charter and in the Universal Declaration of Human Rights.<sup>2</sup> Furthermore, by seeking to assist law enforcement agencies worldwide in leveraging AI to combat online child sexual exploitation and abuse, the initiative engages with what can be considered highly sensitive subjects, namely the use of AI in law enforcement and child sexual exploitation and abuse – a pervasive yet largely unreported violation of children’s rights.

UNICRI and the Ministry of Interior of the UAE are thus taking appropriate measures and setting in place the necessary processes to ensure that the *AI for Safer Children* is ethically and legally developed from the outset. Early internal discussions and analysis identified the main considerations to account for and outlined a plan of action. The decision was taken to conduct further research and prepare an Ethical and Legal Report covering [\(i\) the fundamental principles of the AI for Safer Children initiative](#), [\(ii\) the deriving ethical and legal considerations](#) to account for in the initiative’s implementation, and [\(iii\) the measures that may be taken to address them](#).

This document corresponds to the first version (V1.0) of such a report, intended for discussion in a virtual Ethics Expert Meeting taking place on 7 December 2021 with a group of independent ethics experts to gather their insights. The experts’ feedback will be considered in the preparation of a second version, which will guide UNICRI and the Ministry of Interior of the UAE in the implementation of the initiative. If needed, subsequent versions of the report will be prepared and discussed in follow-up Ethics Expert Meetings.

The considerations flagged in this document and the proposed measures, as well as their exact configuration, are thus not set in stone and will be expanded and improved with the external views collected during and following the Ethics Expert Meeting. At the same time, the process of identifying and addressing the ethical and legal considerations of the *AI for Safer Children* is an ongoing process that will be ever evolving throughout the initiative. This means that the ethical and legal report is not a one-off deliverable. On the contrary, any lessons learned, supervening relevant guidelines or regulatory frameworks or related work from UNICRI<sup>3</sup> will feed into this process.

---

<sup>1</sup> In the absence of a definitive consensus on the definitions of online child sexual exploitation or abuse from the perspective of international law, these two terms are referred to in aggregate to cover all associated concepts. For an explanation of these concepts, see, for instance, the *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse* (ECPAT International 2016) and *Victims Are Not Virtual: Situation Assessment of Online Child Sexual Exploitation In South Asia* (UNICEF 2016).

<sup>2</sup> UN Secretary-General’s Strategy on New Technologies (2018), 4.

<sup>3</sup> Of particular relevance is the upcoming *Toolkit for the Responsible AI Innovation in Law Enforcement* – a joint project of the INTERPOL Innovation Centre and the Centre for AI and Robotics at UNICRI, funded by the European Union.

## I. THE AI FOR SAFER CHILDREN INITIATIVE: AN OVERVIEW

The digital era has enabled unprecedented opportunities to realize children's rights, but it has also intensified children's exposure to risks and harm, including online child sexual exploitation and abuse. While the full scale of online child sexual exploitation and abuse remains unknown, existing evidence suggests considerable cause for concern.<sup>4</sup> The COVID-19 pandemic has had a significant impact on the vulnerability of children, as both children and sex offenders found themselves confined indoors for extended periods of time. At the same time, technological advancements have not only facilitated the amplification of child sexual abuse material but also driven the emergence of entirely new forms of large-scale abuse facilitated by technology such as online grooming and high-definition live-streaming of abuse.<sup>5</sup>

With investigators in law enforcement grappling with burgeoning caseloads and growing backlogs, attention is increasingly turning to new tools and technologies that can help turn the tide in the fight to protect children from sexual exploitation and abuse online. As one of the definitive emerging technologies of the times, AI is at the very core of this development. The potential of AI to support law enforcement in preventing crimes against children has, in fact, already been seen. For instance, facial recognition technology has already enabled the identification of numerous missing children. National agencies are actively exploring how they can further use machine learning to, for instance, identify child abuse images on confiscated devices or rapidly analyze the vast number of reports of potential child sexual abuse material online to swiftly identify children in real danger.

Despite the potential of AI and the current availability of AI tools on the market, law enforcement agencies are not yet widely using AI to combat online child sexual exploitation and abuse. Many law enforcement agencies are not aware of the potential of AI solutions or are sceptical about the technology, due *inter alia* to the sensitivity of the topic of child sexual exploitation and abuse. A further impediment is that selecting and implementing AI tools is a complex process for which law enforcement agencies often lack the necessary knowledge and resources. In fact, AI presents a plethora of challenges from not only a technical perspective, but also from a legal, ethical and societal perspective which must be addressed before the potential of the technology can be realized.

This reality led the UNICRI Centre for AI and Robotics and the Ministry of Interior of the UAE to launch the *AI for Safer Children* initiative in 2020, with the aim of tackling child sexual exploitation and abuse online through the exploration of new technological solutions, specifically AI.

### ***The Two Work Streams of the AI for Safer Children Initiative***

The *AI for Safer Children* initiative consists of two work streams. The first is awareness-raising and outreach, which involves organizing and participating in events for visibility, network-building, awareness-raising and advocacy on the issue of online child sexual exploitation and abuse and how AI can help law enforcement worldwide to tackle it.

---

<sup>4</sup> As reported, for instance, in the 2020 edition of the "Global Report on Trafficking in Persons" of the United Nations Office on Drugs and Crime (UNODC).

<sup>5</sup> Two reports – "Exploiting Isolation: Offenders and victims of online child sexual abuse during the COVID-19 pandemic", released by Europol in June 2020 and "Threats and Trends Child Sexual Exploitation and Abuse: COVID-19 Impact", released by INTERPOL in September 2020 – have noted that, as a result of the pandemic, authorities in several countries have reported an increase in the quantity of child sexual abuse material online and attempts to access websites hosting such content. Furthermore, the reports found that an increase in sexual abuse offences such as online solicitation and sexual extortion has been observed over the course of the pandemic.

The second work stream and the main deliverable of the initiative is the “AI for Safer Children Global Hub” itself, or the “Global Hub” for short. This pillar involves developing and piloting the Global Hub as a platform designed to support law enforcement in leveraging AI to combat online child exploitation and abuse.

The fundamental principles underpinning the *AI for Safer Children* initiative, the ethical and legal considerations identified and the measures intended to address them as outlined in the present Report pertain to these two work streams of the initiative. Yet, before delving into the substance of this Ethical and Legal Report, the central role of the Global Hub justifies a brief expansion upon its design and purpose.

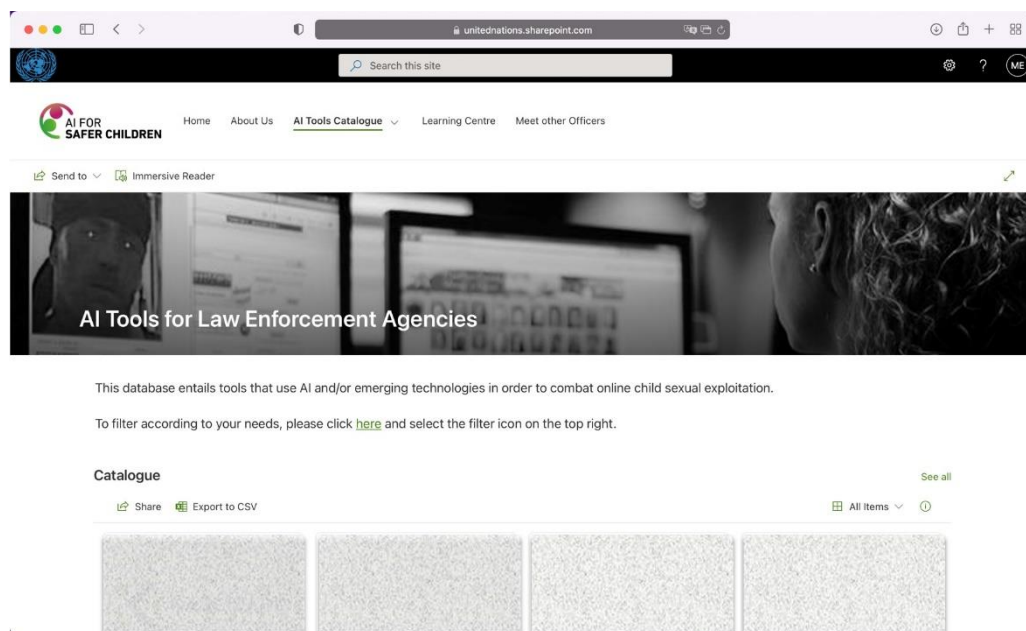
### ***A Closer Look at the AI for Safer Children Global Hub***

The target users of the Global Hub are law enforcement agents from any UN Member State interested in or working with AI tools for combating online child sexual exploitation and abuse. The Global Hub is not an open-source website; rather, these law enforcement users will have exclusive access to the platform via a secure log-in.

The Global Hub pilot platform is envisioned to initially have four core features:

- 1) A *database* of AI tools to tackle online child sexual exploitation and abuse;
- 2) A *selection feature* to help choose the appropriate AI tool for specific contexts and needs;
- 3) *Learning resources* for law enforcement agencies about AI tools; and
- 4) A communication and *networking platform*.

The database is the foundation of the Global Hub, providing a list of existing AI tools – or AI tools under development – that can be leveraged to combat online child sexual exploitation and abuse as well as the ability for law enforcement agents to read and provide feedback about each tool’s specifications. These tools are developed by private companies, academia and not-for-profit organizations.



**Figure 1: The AI database of the AI for Safer Children Global Hub**

The selection feature supports law enforcement users in selecting an appropriate AI tool that meets their specific needs by way of a set of parameters that law enforcement users can select and/or a sequence of questions through which relevant results are provided in a shortlist that can be sorted by users.



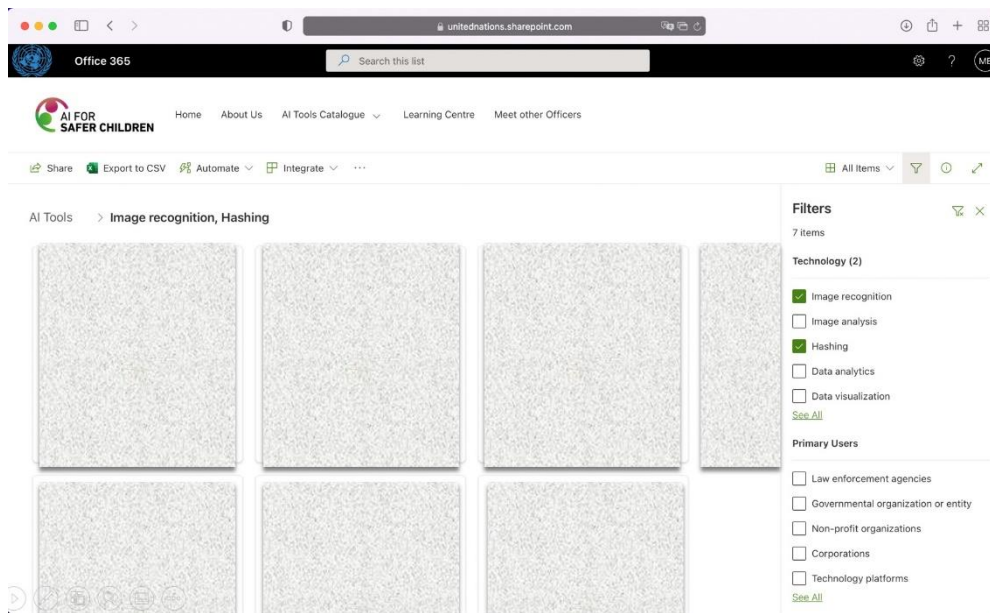


Figure 2: The selection feature of the AI for Safer Children

The third core feature of the platform is intended to build the knowledge of the law enforcement community on AI and how it can be deployed responsibly to support their work. Despite the potential of AI and the existence of AI tools on the market, law enforcement agencies are not yet widely using AI to combat online child sexual exploitation and abuse. And since the use of AI technology is still largely new in the broader law enforcement context, the agents' expertise on the topic is often undeveloped. The learning resources will include, *inter alia*, conceptual explanations, self-paced learning videos, links to training materials available on technology providers' websites, information about upcoming in-person and online trainings and additional resources such as instructions on how to access important data from the National Center for Missing & Exploited Children (NCMEC).

The fourth core feature of the Global Hub, a networking platform, is intended to enable law enforcement users to connect with each other – subject to their consent to sharing contact details for this purpose – and exchange knowledge about AI tools and best practices in the context of child sexual exploitation and abuse online. To help build an active community of practitioners, this feature consists of a contact list of all law enforcement users accessing the platform as well as a section highlighting upcoming events.

## II. THE ETHICAL AND LEGAL PROCESS OF THE AI FOR SAFER CHILDREN INITIATIVE

By aiming to raise awareness and knowledge on how AI can support law enforcement globally to combat online child sexual exploitation and abuse and build the capacity of law enforcement from potentially all UN Member States to adopt currently existing AI tools in this context, the *AI for Safer Children* initiative touches upon what can be considered highly sensitive subjects. Namely, both the use of a new and complex technology such as AI in the law enforcement context and child sexual exploitation and abuse – a pervasive yet largely unreported violation of children's rights – are prominent and sensitive topics which demand full

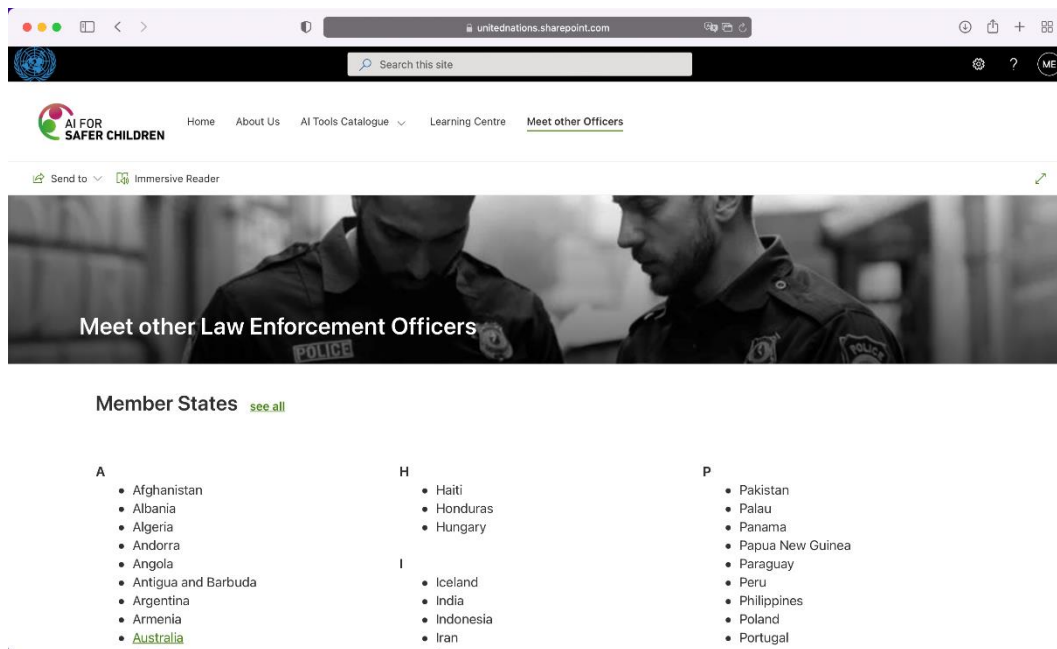


Figure 3: The networking platform of the AI for Safer Children Global

consideration of their technical, legal, ethical and societal complexity; a complexity which substantially increases when looked at from a global perspective.

The *AI for Safer Children* initiative indeed takes a global perspective. It is co-implemented by UNICRI, a UN entity, meaning that it needs to be guided by and fully compliant with the values and principles enshrined in the UN Charter. Article 1 of the Charter established that one of the core purposes of the UN is to promote and encourage respect for human rights and fundamental freedoms.

These reasons lead UNICRI and Ministry of Interior of the UAE to establish an ethical and legal process of the *AI for Safer Children* initiative that has been ongoing from the start of the initiative, in parallel to its other activities such as the design of the Global Hub, engagement with stakeholders and participation and organization of outreach events. Although the term “ethics” is often used in a broad way to encompass also legal considerations such as those emergent from human rights law, the process (as well as the current report) has been entitled “ethical and legal” to highlight that there are also legal obligations at stake, including those deriving from the UN Charter and international human rights law.

So far, the ethical and legal process has comprised of internal coordination meetings dedicated to this topic, collecting of feedback during the AI for Safer Children stakeholder meetings on 15 – 17 June 2021, as well as extensive research and consultations to prepare the present first version of the *AI for Safer Children* Ethical and Legal Report.

The research involved in drafting the Ethical and Legal Report involved desk-based research for legal and ethical frameworks and high-level guidance both applicable to and relevant for the initiative, as well as for precedents from similar initiatives. Adding to this, numerous dedicated interviews and meetings with ethics experts, law enforcement agents, representatives of technology providers and colleagues from other international organizations grappling with similar situations, both in and outside the UN, were undertaken to gather different perspectives and feedback on the soundness of the initiative’s ethical and legal process and the best strategies to adopt therein. The next sections are the result of this research and these interviews.

Upcoming milestones of the ethical and legal process include the Ethics Expert Meeting on 7 December 2021, where ethics experts and other relevant participants will provide inputs and analysis. The relevant



feedback from this meeting will be implemented in the second version of the report. If deemed necessary, further Ethics Expert Meetings or individualized interviews will be convened to gather views of the second or subsequent versions of the report or on particular considerations and related measures.

In parallel, the measures outlined in the report will be implemented to the extent allowed by the scope of the *AI for Safer Children* initiative. These are not necessarily one-off solutions and may include the establishment of ongoing mechanisms and procedures to address certain ethical issues or identify possible future issues which could be addressed and implemented in the post-pilot phase.

### III. THE CORE PRINCIPLES OF THE *AI FOR SAFER CHILDREN* INITIATIVE

To ensure the ethical and legal soundness of the *AI for Safer Children* initiative, the principles guiding its development must first be set out.

Four core principles underpinning the initiative are highlighted below, which constitute the bases for the ensuing ethical and legal considerations analyzed in the following section.

- 1 *The values enshrined in both the UN Charter and the Universal Declaration of Human Rights and human rights principles at large are at the forefront of the AI for Safer Children initiative.*

As stated in the very first article of the UN Charter, one of the core principles of the UN is to promote and encourage respect for human rights and fundamental freedoms. International human rights law, moreover, provides a universally agreed-upon set of principles on which to build the *AI for Safer Children* initiative's promotion and encouragement of the responsible use of AI technology, including the human rights principles enshrined in the Universal Declaration of Human Rights. This framework was distinguished as a necessary underpinning of the entire UN system by the UN Secretary-General in his Call to Action for Human Rights in 2020.

- 2 *The AI for Safer Children initiative will foster the responsible development and use of AI in law enforcement, in line with UN values.*

As with all projects related to new and emerging technologies, the *AI for Safer Children* initiative must be wary of promoting the responsible development and use of AI. This aspect is crucial to any UN-related activity in order to uphold universal human rights standards noted in Principle 2 below. As negative consequences involving AI tools may stem either from the development process of the tool or as a result of the manner of its use, the initiative must promote the responsible design and use of AI in line with the UN values, respect for internationally recognized human rights law, the rule of law, and the principle of democracy. As data is a fundamental resource in the development of AI, universal principles and requirement of data protection require special consideration.

- 3 *The AI for Safer Children initiative promotes global inclusion targeting all UN Member States and gathers perspectives from a diverse range of stakeholders, including those at risk of being left furthest behind.*

Human rights can only be fully realized when linked to another fundamental value of the UN – inclusion of all peoples and stakeholder groups. Online child sexual exploitation and abuse is a global issue which calls for special consideration of the needs of those most at risk of being left behind by the digital divide, such as women and girls and people in the Global South. Digital cooperation is therefore essential in combating the exacerbation of existing inequalities. Moreover, as a project implemented by a UN entity, the *AI for Safer Children* initiative should benefit all Member States equally. The initiative should therefore strive to engage in meaningful dialogue with a diverse range of perspectives from law enforcement authorities, governments, civil society, businesses and other stakeholders across the globe – including those representing people at risk of being left furthest behind, particularly women and girls.

**4** *The values of integrity, impartiality, independence and transparency will be upheld in all aspects of the development of the AI for Safer Children initiative.*

The *AI for Safer Children* initiative adheres to the highest standards of integrity and impartiality in pursuing its goals, as it is guided by existing norms and standards of international law. To that end, the initiative must also be as transparent as possible about its scope and activities.

#### **IV. FROM PRINCIPLES TO ACTIONS: THE ETHICAL AND LEGAL CONSIDERATIONS AND HOW TO ADDRESS THEM**

Bearing in mind the principles highlighted in the previous section, this section expands on specific ethical and legal considerations that were identified as particularly relevant in the implementation of the *AI for Safer Children* initiative and specifies the measures that are envisioned and/or already being implemented to address each consideration.

##### **1. Promoting Global Inclusiveness in a Digitally Divided World**

Above all, the goal of the initiative is to support law enforcement worldwide in responsibly using AI to combat online child sexual exploitation and abuse, from which all UN Members States can benefit. The growing digital divide necessitates that the initiative be particularly mindful of those countries at risk of being left behind. This means supporting different stakeholders and being open and available for global engagement at every opportunity throughout the development of the initiative.

Law enforcement agencies across the world must feel that their concerns are respected and that the initiative is suited to their needs, resources and capabilities. By thus including countries at risk of being left behind by the digital divide in discussions throughout its development and taking into account their perspectives, the initiative aims to gather further recommendations for how to best reach these countries.

##### **Concrete measures addressing this consideration include:**

- Raising awareness about the issue of online child sexual exploitation and abuse and its global nature along with the limitations caused by the digital divide, notably the reduced amounts of data available for certain countries and regions, particularly in the Global South.
- Ensuring geographical diversity of the participants and speakers at the events.
- Promoting the initiative globally and taking into account the perspectives and needs of law enforcement agents from countries across all regions in the development of the initiative, particularly from those at risk of being left behind.
- In the design of the Global Hub, writing the texts in accessible language, e.g. avoiding over-complicated phrasing or unnecessarily technical language, and translating subsequent versions of the platform from English into at least other official UN languages.
- Including a learning section on the Global Hub to build the capacity of law enforcement agents worldwide with different levels of technical proficiency to responsibly use AI in their work.
- Researching AI tools developed in different countries across all regions to be featured on the Global Hub.
- Adding, in the future, a national and regional training component to the initiative.
- Collecting information from the technology providers of the AI tools to be featured on the Global Hub about which languages the tool is available in and, for those with text analysis capabilities, which languages the tool can analyze.

##### **2. Gender Mainstreaming**

The digital divide has exacerbated existing issues of gender inequality, leading to fewer women and girls participating in both the technology industry and technological development in general when now is a critical

time to include their specific needs and perspectives for them to benefit from the opportunities of society's digital transformation. While human rights underpin the work of all UN activities, these will never be realized without a special emphasis on the human rights of women and girls – one of the demographics left furthest behind by the digital divide. Therefore, it is particularly important that tech-centred initiatives such as *AI for Safer Children* set women and girls' considerations at the centre of their processes, along with others most at risk of being left behind.

**Concrete measures addressing this consideration include:**

- Raising awareness about the risks of gender biases in the development of the AI tools (for instance, because of under-representation of women and girls in the training data), with the consequent negative impact on women and girls if used by law enforcement without regard to this bias. This would be achieved by including the information on the Global Hub's learning section as well as highlighting it in advocacy and outreach events.
- Encouraging law enforcement agencies to ensure gender balance of the law enforcement agents they designate as their representatives on the Global Hub.
- Ensuring gender diversity of the participants and speakers at meetings and events.  
Ensuring gender diversity of the team developing the *AI for Safer Children* initiative.

### **3. Protection of Personal Data Processed Within the Initiative**

Regarding the right to privacy, any personal data collected in relation to the *AI for Safer Children* initiative must be protected and only shared with the consent of the data subjects and processed according to the UN Personal Data Protection and Privacy Principles. Thus, the data should be processed in a manner that is fair, limited and adequate to what is necessary for specified purposes of the initiative and consistent with UN's mandate.

Protecting the personal data collected is of fundamental importance for trust in the initiative. All law enforcement agents using the AI for Safer Children Global Hub, speakers and participants of events, etc., are entitled to the right to privacy and can only share freely if they trust that their data will be used in a responsible manner for the collective, positive outcome envisioned by the initiative.

The initiative must therefore have a transparent data processing policy which only retains personal data for the time that is necessary for the specified purposes and must be processed with due regard to confidentiality and accuracy. The collected information will be stored on official, secure servers as another measure to protect the security and confidentiality of the data.

**Concrete measures addressing this consideration include:**

- Implementing and following a transparent data processing policy in line with the UN Personal Data Protection and Privacy Principles.
- Utilizing an official, secure server on which to store collected data.

### **4. Security of the AI for Safer Children Global Hub**

Guided by the principle of integrity and fostering a sense of trust in both the law enforcement users and the broader community, the Global Hub needs to be a secure platform, protected against any unauthorized access, attacks, disruptions, abuses or any other risks to the safety of the information included therein.

The cybersecurity of the Global Hub also relates to the previous consideration on personal data protection, as it safeguards any personal information of the law enforcement users, thereby protecting their right to privacy and adhering to the UN Data Protection and Privacy Principles. Furthermore, ensuring cybersecurity is crucial to protect any information about the AI tools disclosed by the technology providers, encouraging them to share the required information more openly and truthfully.

**Concrete measures addressing this consideration include:**

- Hosting the Global Hub under the supervision of the United Nations Office of Information and Communications Technology (OICT), which has experience with other secure websites and the gambit of possible cybersecurity mechanisms.
- Limiting access to the Global Hub to law enforcement agents through, for instance, a secure log-in and possibly requiring further verification of their identity and affiliation.

**5. Avoid featuring unethical or unlawful tools on the Global Hub**

Linked with the core purpose of the *AI for Safer Children* initiative to provide information to law enforcement agents worldwide on existing AI tools that can help combating child sexual exploitation and abuse online is the need to prevent the Global Hub from featuring any tools that raise serious ethical and legal concerns – such as tools whose developers fail to take the necessary corrective or mitigating measures to prevent grave human rights violations resulting from use or misuse of the tools. A consideration particular to the initiative is that the tools are developed with due regard for children's rights as covered in the 1989 Convention on the Rights of the Child, ensuring that their personal data is handled with care and any potential harmful effects to either child victims or other children incidentally involved are duly researched and considered.<sup>6</sup>

The AI tools which the Global Hub will contain are tools that are being developed or were already placed on the market by the private industry, academia or non-profit organizations independently from the *AI for Safer Children* initiative. Moreover, it is outside of the scope of the initiative to evaluate, test or label certain AI tools as ethical or unethical. Nonetheless, part of being guided by human rights principles and the aim to promote responsible AI innovation in law enforcement involves that the initiative foster an environment emphasizing ethical AI development as well as avoiding endorsement of any tools that are problematically developed from a legal and ethical standpoint.

To avoid featuring unethical or unlawful tools on the Global Hub, the *AI for Safer Children* initiative has invested research into the possible measures that are within the initiative's scope and at the same time do not discourage technology providers from participating in the initiative.

**Concrete measures addressing this consideration include:**

- Being clear and transparent about the core principles and scope of the AI for Safer Children initiative, both by providing a comprehensive statement as such on the Global Hub and by making publicly available documents such as the present Ethical and Legal Report.
- Collecting information from the technology providers on the ethical considerations involved in their development of the AI tools featured on the Global Hub and providing this information to the law enforcement users.
- Having processes in place to ensure that, if reliable information that any tool raises ethical and legal issues is received, such tools are removed from the Global Hub and its users are informed.

Due to the involved nature of these measures, they are further elaborated below.

***Statement Document for Transparency***

In the interest of transparency and promoting the responsible development and use of AI in line with human rights, a statement document should be produced and made available on the Global Hub which clearly communicates the core principles underpinning the *AI for Safer Children* initiative and its scope.

---

<sup>6</sup> Guidelines can be drawn from UNICEF's "[Policy Guidance on AI for Children](#)", a draft version of which was published in 2020, and UNESCO's "[Recommendation on the ethics of artificial intelligence](#)", from 2021.

While the specific content of such a document is not yet defined, it should make clear as a form of disclaimer that the AI for Safer Children is an initiative simply providing information; that it is not an assessment of the tools and that their inclusion in the database does not equal endorsement by the UN since the initiative does not involve testing whether the principles of responsible AI are upheld by the AI tools. The statement document should also communicate any measures being taken within the initiative to nevertheless avoid featuring unethical or unlawful tools in the database.

Making this statement available to technology providers has the added value of raising awareness about and drawing attention to the importance of responsibly developing AI as a prerequisite for its use to have the desired positive impact.

### ***Questionnaire for Technology Providers on the Ethical Aspects of the Tools' Development***

Another measure addressing the consideration of how to avoid featuring unethical or unlawful tools on the Global Hub would be to collect information from the providers on the ethical aspects of the development of their tools which they addressed during the development phase. These may include, for instance, whether the training data is representative of different genders and disadvantaged groups, whether the team developing the tools is diverse, which personal data protection measures were taken, whether special consideration was given to children's rights and well-being, and which measures were taken to ensure the robustness and cybersecurity of the tools.<sup>7</sup> Providers who want to have their tools featured on the Global Hub are already being requested to complete a questionnaire on the technical specifications of the tools. Providers could then be requested to answer another set of questions regarding the ethical considerations of the tools' development.

The questionnaire is not an assessment of the AI tools, as, again, that would fall outside the scope of the initiative. Its purpose is instead to raise awareness among law enforcement users about the importance of ethics, as much as technical efficiency, being involved in the development of the AI tool for it to ultimately be effective in supporting their needs.

The contents and language (*i.e.* the framing of the questions and how to communicate the purpose of the questionnaire in the most productive manner) of this ethics questionnaire require further research and consultation, including with representatives from law enforcement and technology providers, before being specified. However, preliminary feedback received from a few law enforcement agents about the type of information in which they would be interested indicated that the following areas would be the most relevant to them:

- The strengths and weaknesses of the tools, to understand how and where to implement it for the best results (which would be a positive outcome for both law enforcement and the technology provider). Such information would involve questions about which datasets the algorithm was trained on, the ethical/human rights risk detection and mitigating measures taken by the developers, and the accuracy and limitations of their tool.
- Whether the tool has been independently verified as accurate or fair, or been successfully (or otherwise) tested and used by another law enforcement agency. Another factor is who or what the AI tool was originally developed for, and for what purpose.

---

<sup>7</sup> When devising the information sought in the questionnaire about the developers' ethical processes, guidance will be drawn from authoritative global sources about what considerations are most pertinent. These sources include, e.g., the *Recommendation on the Ethics of Artificial Intelligence*, adopted on 24 November 2021 by UNESCO's General Conference at its 41<sup>st</sup> session, and UNICEF's *Policy guidance on AI for children*.

Other ongoing work from UNICRI will also feed into this process, particularly the upcoming INTERPOL and UNICRI *Toolkit for the Responsible AI Innovation in Law Enforcement* – a joint project of the INTERPOL Innovation Centre and the Centre for AI and Robotics at UNICRI, funded by the European Union.



- Whether the AI tool collects any user data (or data analytics) from the law enforcement agents and, if so, what data and for what purposes.

Likewise, the questionnaire aims to raise awareness among the technology providers about the type of concerns that law enforcement users may have to face regarding the ethical side of the technology's development before adopting it. In this regard, it would benefit the providers to be forthcoming about what ethical considerations were involved in the development of their tool, although it is up to each developer to fill in the questionnaire as they see fit and up to each user how they make use of the information provided therein.

The incentives, and how to frame them, for developers to fill out the questionnaire fully and conscientiously are open for discussion; as are the details of the questionnaire's design – e.g., how long it should be, whether questions are best open-ended or not, how detailed, etc. These concerns are pending input from experts and stakeholders and further research.

### ***Removing Unethical or Unlawful Tools from the Global Hub***

A further measure envisioned in this context is instituting a process for removing any AI tool that is found to raise serious ethical and legal concerns. Although the *AI for Safer Children* initiative will not involve individual testing or assessment of the AI tools, the principles underlying the initiative require that the integrity of the database is upheld by expelling a tool if any such concerns come to light, whether through the communication platform or otherwise.

Therefore, where reliable information is received leading to the substantiated conclusion that a given AI tool raises serious concerns, that tool must be removed from the Global Hub. If feasible, guidance and support could be given to law enforcement users of the tool in question about how to address possible concerns, as well as to technology providers about what specific measure in the development process may mitigate similar concerns. For transparency purposes, and to emphasize the principles shaping the initiative, this measure shall be communicated to both the technology providers and the law enforcement users.

## **6. Promote the Responsible Use of AI by Law Enforcement**

As an initiative which seeks to bring knowledge about AI tools to law enforcement agencies worldwide combating online child sexual exploitation and abuse, the *AI for Safer Children* initiative must promote the use of AI in a responsible manner in all of its awareness-raising activities as well as on the Global Hub. . While it is beyond the scope of the *AI for Safer Children* initiative to be involved in each country's implementation of the AI tools, the initiative should nevertheless build the knowledge and understanding of the law enforcement users on how they can use these tools responsibly.

Two of the concrete measures already highlighted in the previous section also aim to address this consideration.

- Firstly, providing the information collected from the technology providers about the ethics behind their AI tools to the law enforcement users. Making the questionnaire completed by the technology providers available next to each tool would ideally serve to promote consideration of the issues and questions therein – alerting law enforcement to the necessity of considering, not just the efficiency yet tied to it, the ethical concerns involved in using AI technology. For instance, the robust and secure processing mechanisms for personal data involving children require special consideration in light of children's rights.
- Secondly, being clear and transparent about the core principles and scope of the AI for Safer Children initiative. By presenting the statement document mentioned in the previous section to the law enforcement users is a way of raising awareness with them about how such principles must always be considered in both the development and the use of AI in law enforcement.

In addition to these measures, **the following concrete measure was identified:**

- Including guidance on the Global Hub's learning section specifically dedicated to the responsible use of AI, from developing a tool in-house to the considerations of its integration into the workforce and deployment.

The design of this learning section is further elaborated below.

### ***Learning Section***

To promote the responsible use of AI tools, the Global Hub will include content in its learning section specifically dedicated to the responsible use of AI in law enforcement. It is envisioned to include content such as videos and tutorials about the use of AI and links to external AI ethics guidelines and resources, such as the *Recommendation on the Ethics of Artificial Intelligence*, adopted on 24 November 2021 by UNESCO's General Conference at its 41st session, and UNICEF's *Policy guidance on AI for children*. A more developed solution would be to include on the Global Hub a tailored and summarized version of the upcoming UNICRI-INTERPOL *Toolkit for Responsible AI Innovation in Law Enforcement*<sup>8</sup> with specific guidance on how principles of responsible AI translate to practical implementation of AI tools in law enforcement.

### **7. Fair Promotion of AI providers**

Related to principles of human rights, global inclusivity and integrity, bigger AI providers or those associated with the Advisory Board of the *AI for Safer Children* shall not be favoured. AI providers with fewer resources should be given the same chances to participate in the *AI for Safer Children* initiative and have their AI tools included on the Global Hub as AI providers with more resources. The awareness-raising stream of the *AI for Safer Children* initiative should also keep this principle of fair promotion of AI providers under consideration, for instance in the selection of speakers from the technology industry and as participants at events and activities related to the initiative.

This fairness has the added benefit of ensuring that law enforcement agencies can identify the AI tool that best suits their needs according to the criteria they input, rather than being influenced in favour of certain providers for any other reason.

**Concrete measures addressing this consideration include:**

- Ensuring that the Global Hub does not favour any AI tool, for instance by giving every tool equal visibility on the platform and not allowing any kind of commercial advertisement.
- Maintaining an impartial relationship with technology providers by not favouring any specific provider as a speaker or participant at events and activities related to the initiative.
- Proactively researching the AI tools to be featured on the Global Hub instead of relying on private industry to reach out to the initiative, thereby preventing bigger technology providers from being overrepresented.

### **Conclusion**

The present report has outlined how the ethical and legal process of the *AI for Safer Children* initiative identified core principles and ensuing concerns regarding its activities, and the concrete actions it has taken to practically instil those principles and address the ethical concerns to ultimately contribute to the best possible outcome: supporting law enforcement globally to responsibly use AI to aid in combating online child sexual exploitation and abuse.

---

<sup>8</sup> A joint project of the INTERPOL Innovation Centre and the Centre for AI and Robotics at UNICRI, funded by the European Union.