INTERPOL

unicri
United Nations
Interregional Crime and Justice
Research Institute

RESPONSIBLE
AI INNOVATION IN
LAW ENFORCEMENT

AI Toolkit

Risk Assessment Questionnaire

Funded by
the European Union

REVISED FEBRUARY 2024

INTERPOL

United Nations
Interregional Crime and Justice
Research Institute

Funded by
the European Union

DISCLAIMER

The contents of this document are for information purposes only. INTERPOL and UNICRI assume no liability or responsibility for any inaccurate or incomplete information, nor for any actions taken in reliance thereon. The published material is distributed without warranty of any kind, either express or implied, and the responsibility for the interpretation and use of the material lies with the reader. In no event shall, INTERPOL or UNICRI be liable for damages arising from its use.

INTERPOL and UNICRI take no responsibility for the content of any external website referenced in this publication or for any defamatory, offensive or misleading information which might be contained on these third-party websites. Any links to external websites do not constitute an endorsement by INTERPOL or UNICRI, and are only provided as a convenience. It is the responsibility of the reader to evaluate the content and usefulness of information obtained from other sites.

The views, thoughts and opinions expressed in the content of this publication belong solely to the authors and do not necessarily reflect the views or policies of INTERPOL or the United Nations, their member countries or member states, their governing bodies, or contributory organizations, nor does it imply any endorsement. Therefore, INTERPOL and UNICRI carry no responsibility for the opinions expressed in this publication.

INTERPOL and UNICRI do not endorse or recommend any product, process, or service. Therefore, mention of any products, processes, or services in this document cannot be construed as an endorsement or recommendation by INTERPOL or UNICRI.

The designation employed and presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations, UNICRI or INTERPOL, concerning the legal status of any country, territory, city or area of its authorities, or concerning the delimitation of its frontiers or boundaries. The contents of this document may be quoted or reproduced, provided that the source of information is acknowledged. INTERPOL and UNICRI would like to receive a copy of the document in which this publication is used or quoted.

OVERVIEW

WHAT

The Risk Assessment Questionnaire is intended to support law enforcement agencies to evaluate the risks an AI system may pose, from a responsible AI innovation perspective. More specifically, it supports law enforcement agencies with identifying the potential adverse impacts on individuals, groups, and society as a whole, as well as the probability of such impacts occurring. An important part of the Risk Assessment Questionnaire is identifying and assessing the organizational, technical and security measures established as part of the design, development, procurement, deployment or use of the AI system. This will help agencies to assess the likelihood of unintended impacts, highlight risk levels, and provide an overview of the risks connected to the use of the AI system.

WHEN

It is recommended that this Risk Assessment Questionnaire should be used early in the design and development phases and when testing and evaluating an AI system. Beyond these defined periods, agencies may also benefit from running the Risk Assessment Questionnaire periodically over the course of an AI system's life cycle, in order to make the best possible use of the insights it provides. For instance, the Risk Assessment Questionnaire can be helpful for decision-makers when determining whether to approve, change, or decommission an AI system, when handling residual risks, or as a tool to raise awareness among users of the adverse impacts of an AI system.

WHO

The Risk Assessment Questionnaire is designed to be completed by the team or staff member(s) within the law enforcement agency who are in charge of the overall AI project or a specific stage within the AI life cycle.

TABLE OF CONTENTS

Before you start

Practicing responsible AI Innovation involves identifying and understanding the potential weaknesses and threats that may emerge during the AI innovation journey. It also involves taking informed action to prevent and mitigate such weaknesses and threats, all in accordance with the ***Principles for Responsible AI Innovation.*** These activities should be part of a law enforcement agency's broader risk management process, which should be defined and implemented at an organization-wide level. |► *Learn more about risk management and its importance in the* ***Organizational Roadmap***. An accurate assessment of the risks of the AI system that the law enforcement agency intends to use is a crucial step in this process.

This Risk Assessment Questionnaire will support you and your agency in this regard by aiding you to detect and evaluate the risks to individuals and communities that may emerge over the course of the life cycle of an AI system that is used or intended for use in your agency.

Conducting this Risk Assessment Questionnaire can be considered a precondition for agencies if they are to appropriately and adequately develop strategies to prevent and mitigate the risks posed by an AI system; identify and implement the responses to such risks, including mitigation measures; and prioritize limited resources when responding to situations with different risk levels.

The scope of this risk assessment questionnaire

While there are different kinds of risk assessment that cover various types of risk – such as information and cybersecurity risks, reputational and financial risks, and data protection or human rights risks – **this Risk Assessment Questionnaire focuses exclusively on the negative consequences for individuals and communities should the *Principles for Responsible AI Innovation* not be adhered to for any particular AI system used or intended for use by an agency.** This Risk Assessment Questionnaire is not meant to replace any of these other kinds of risk assessment; neither has it been designed to incorporate these other kinds of assessment. Rather, it is intended to complement existing risk assessment practices that law enforcement agencies already carry out or are required to carry out under the applicable law.

Instructions

This Risk Assessment contains questions that concern:

- the likelihood of certain negative events or circumstances occurring.
- the impact of such events or circumstances on individuals and communities should they occur.

These questions draw on the ***Principles for Responsible AI Innovation,*** particularly the core principles of minimization of harm, human autonomy, and fairness, and the corresponding instrumental principles. While there are no questions directly focusing on the principle of good governance, this principle is addressed on two separate occasions. *First,* the extent to which law enforcement agencies follow the principle of good governance may influence the likelihood and/or the impact of a certain event. For example, ensuring the traceability of the decisions taken during the design and development of the AI system increases the chances of identifying harmful human biases in the system, thus allowing for their early prevention. *Second,* the principle of good governance guides agencies and personnel in the identification and implementation of mitigation measures.

The process of undergoing this Risk Assessment Questionnaire includes six steps, all of which are interconnected and will contribute to the overall effectiveness of the assessment. These steps include:

1. Preparing for the risk assessment
2. Completing the questionnaire
3. Interpreting the results
4. Communicating the results to the relevant risk owner
5. Responding to the risks; and
6. Repeating the questionnaire.[1]

STEP 1: PREPARING FOR THE QUESTIONNAIRE

Prior to conducting the Risk Assessment Questionnaire, it is important to first examine and fully understand the situation as a whole. To this end, you should endeavour to prepare answers to the questions listed below. It is recommended that internal or external legal experts be involved in this exercise.

1. At what stage of the AI life cycle will the Risk Assessment Questionnaire be completed? What timeframe will it cover?

2. What is the AI system use case to be assessed?

3. What are the potential limitations of the Risk Assessment Questionnaire? Which other types of risk assessments could provide additional information?

4. What sources of information will you use to fill in the Risk Assessment Questionnaire? Are they sufficient or do you need access to certain other information? How can you gain access to this information?

5. What are the foreseen or observed weaknesses regarding the implementation of the ***Principles for Responsible AI Innovation?*** *Refer to the results of the "Principles in Action" exercise(s) in the* ***Responsible AI Innovation in Action Workbook.***

6. What are the stakeholders of the AI system's implementation? *Refer to the results of the "Stakeholder Engagement" exercise(s) in the* ***Responsible AI Innovation in Action Workbook***.

7. What is the context and mode of acquisition for the AI system? *Refer to the results of the "Gaps & Needs Analysis", "Value Mapping" and "Deciding how to obtain the AI System" exercises in the Planning State of the* ***Responsible AI Innovation in Action Workbook***.

8. How could the AI system be misused or accessed by unauthorized users and what would be the consequences of that? *Refer to the results of the "Identifying possible misuses or unauthorized uses" exercise in the* ***Responsible AI Innovation in Action Workbook*** *and any information and cybersecurity risk assessment or similar risk assessments carried out in our agency.*

9. What possible negative impacts on human rights may result from the use of the AI system? *Refer to the results of the human rights impact assessment carried out in relation to the AI system.*

STEP 2: FILLING OUT THE QUESTIONNAIRE

To conduct the Risk Assessment, the team or staff member(s) within the law enforcement agency who are in charge of the AI project or a specific stage within the AI life cycle should complete the questionnaire below. This may require consultation with other relevant internal or external parties who have the necessary knowledge and understanding of the potential risks of the AI system.

The Risk Assessment Questionnaire is structured around two main categories: impact and likelihood. A score of 1 to 5 should be assigned to each of the questions provided under each of these two categories. Once all the questions have been answered and scored, the respondent will be able to calculate the overall risk score by multiplying the impact score by the likelihood score for each risk (Risk Score = Impact x Likelihood). The risk score offers a quantitative measure of the risk level, thus enabling the law enforcement agency to understand specific risks which may undermine the ***Principles for Responsible AI Innovation***.

ANNEX A contains a glossary of key terms used in the Risk Assessment Questionnaire.

STEP 3: INTERPRETING THE RESULTS

Having determined the risk score, a risk level – classified as low, medium, high or extreme risk – can then be determined. The risk level demonstrates the extent or magnitude of a particular risk. Identifying the risk level can be helpful in communicating the results of the self-assessment. The risk matrix below may help with this exercise.

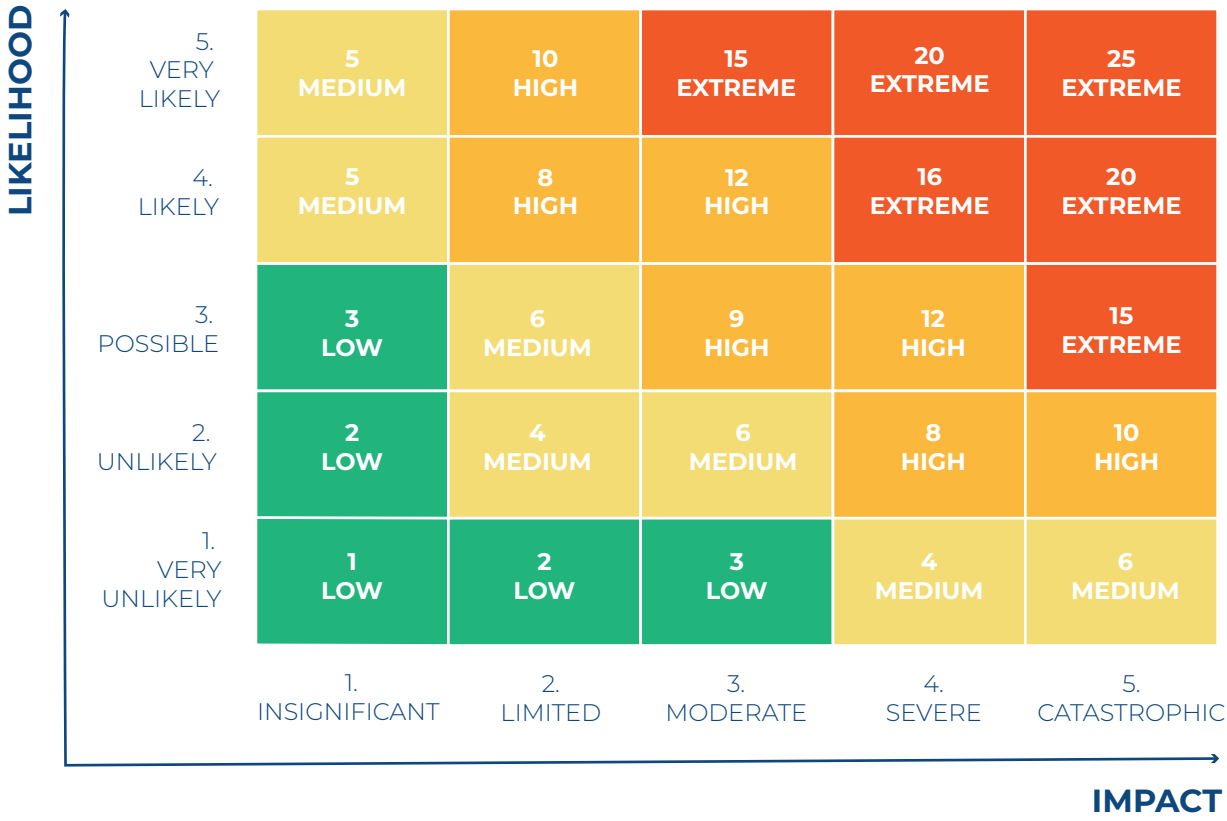

Figure 1 Risk Matrix

The table below indicates the scores that correspond to each risk level and provides a general interpretation of each.

	SCORE RANGE	INTERPRETATION
LOW RISK	1 to 3	The likelihood of the event occurring is very low and/or, if it does occur, the impact on individuals and communities from a human rights and ethics perspective will be minimal. This may include AI systems that produce minor inaccuracies or when its use leads to temporary disruption of non-critical services. Despite being classified as low, these risks should still be managed and prevented or mitigated where possible.
MEDIUM RISK	4 to 6	The likelihood of the event occurring is low, or, if there is a higher likelihood, the impact if it does occur will not be severe. This could involve cases such as unintentional bias in non-critical decision-making processes that do not lead to significant harm. These risks require more active management and planning to mitigate.
HIGH RISK	8 to 12	The event is likely to occur, or the impact if it does occur will be severe. This could include cases such as significant biases in critical decision-making processes or privacy violations that lead to serious harm. These risks require immediate attention and robust mitigation strategies.
EXTREME RISK	15 to 25	The risk is almost certain to occur, or the impact if it does occur will be extremely severe or catastrophic. This could involve cases such as widespread violations of human rights or other serious harm to individuals. These risks require urgent, comprehensive action, including redesigning or discontinuing the AI system.

STEP 4: COMMUNICATING THE RESULTS

Once the results have been obtained and interpreted, they should be conveyed to the relevant risk owner – the individual(s) or team(s) with the authority and accountability to administer and/or respond to the risks. This communication should include the risk score and a detailed explanation of the findings of the assessment, combined with the information gathered during the preparation phase. The risk owner should use all the relevant information to formulate an effective risk mitigation strategy, in collaboration with the individual or team responsible for the AI system at the stage in question.

The parties involved in this step can use the "Risk Response" exercise in the ***Responsible AI Innovation in Action Workbook*** to communicate the results of the Risk Assessment and outline the risk response.

STEP 5: RESPONDING TO THE RISK

WANT TO LEARN MORE?
See Annex B: "Risk Mitigation Measures"

Once a risk has been assessed and identified, the organization and/or the risk owner is prompted to choose a course of action to deal with the risk. When the level of risk is in the Medium, High or Extreme category, it is particularly important that adequate measures are taken to minimize the severity of the risk's impact and/or its likelihood. These measures should be sufficient to lower the risk level to Low and safeguard individuals and communities from continuous and disproportionate harm.

The risk response may include one or more of the following types of measures:

- **Risk acceptance:** Making an informed and conscious decision to take on a specific risk with the aim of leveraging the opportunities offered by the AI system. It is crucial that these accepted risks are in the Low category and are continuously monitored and reviewed to ensure the good governance principles of traceability, auditability and accountability are adhered to.

- **Risk mitigation:** Implementing measures designed to decrease the level of risk. These strategies can include various processes, policies, practices, or devices intended to decrease the probability and/or severity of a particular event or situation, or eliminate the source of the risk.

- **Risk avoidance:** Making a conscious and informed decision to refrain from participating in or to terminate the creation, deployment, or use of an AI system in order to prevent a specific risk. Risk avoidance may be based on the outcomes of a risk assessment and/or legal and regulatory responsibilities.

- **Risk monitoring:** Setting up constant observation, supervision, and evaluation of the risk status to identify any instances where the use of the AI system creates harm and other negative effects on individuals and communities. This may also involve documenting all processes associated with monitoring the AI system in order to

guarantee traceability and auditability. This measure is also valuable as it enables the assessment of the efficacy of any risk mitigation measures previously implemented and supports determining whether new measures need to be introduced.

- **Risk auditing:** Instituting an objective, systematic, and thoroughly documented procedure carried out by a third party in order to independently assess the adequacy and effectiveness of the AI risk management framework in terms of protecting individuals and communities from harm and other negative effects.

- **Risk reporting:** Initiating a communication process in order to inform certain internal or external stakeholders of the current status of the risks associated with the AI system and how they are being managed.

- **Risk financing:** Allocating funds to address or alter any potential financial consequences of the risks identified.

STEP 6: REPEATING THE QUESTIONNAIRE

The risk assessment process for an AI system is iterative. The Risk Assessment Questionnaire should be repeated over time, allowing for the continuous maintenance, monitoring and reevaluation of risks as circumstances evolve. There are two main circumstances under which the Risk Assessment Questionnaire should be repeated:

1. **Risk reassessment post mitigation**: It is recommended that the Risk Assessment should be carried out again after mitigation measures have been implemented. This serves to assess whether these measures have effectively modified the risk level. If the risk level remains high, further measures may need to be taken. If the risk level is reduced, the measures may need to be maintained or adjusted accordingly.

2. **Periodic and needs-based assessment**: Given the highly dynamic nature of AI technology and its broad scope of application, the Risk Assessment Questionnaire process should be conducted periodically and as and when necessary. This is especially crucial when any risk-modifying factors have been identified. These factors can include the integration of new mitigation measures, changes in the technical components of the AI system, alterations in its operating environment, and the introduction of new stakeholders, among others. These changes can significantly impact the risk profile of the AI system, necessitating a reevaluation of the associated risks.

The parties involved in this step can use the “Risk Response” and “Risk Monitoring” exercises in the ***Responsible AI Innovation in Action Workbook*** to help analyze and monitor the response to the AI system’s risks.

Questionnaire

The questions below are organized according to the core principles for responsible AI innovation they refer to, and a short definition of each principle is provided to facilitate understanding of the questions. Each pair of questions includes a list of practical examples of potential harm or negative effects that a circumstance or event would have on individuals and communities. These examples are not exhaustive and simply intend to illustrate the types of impact resulting from implementing the AI system to support you in conducting the assessment.

The order in which the questions below are presented is designed to ensure clarity. However, in practice, it is expected that there will be some back and forth between the questions. It may also be beneficial to begin with the questions about impact before moving on to the questions about likelihood.

LAWFULNESS

LIKELIHOOD (A) 1 – Very Unlikely; 2 – Unlikely; 3 – Possible; 4 – Likely; 5 – Very likely		IMPACT (B) 1 – Insignificant; 2 – Limited; 3 – Moderate; 4 – Severe; 5 - Catastrophic		RISK LEVEL (AxB)	RISK OWNER	OBSERVATIONS
EXAMPLES: Non-compliance with the applicable laws and regulations in activities related to the AI system could, for example, lead to: • Discrimination of individuals who suffer a disproportionate negative impact due to the AI system's use. • A breach of privacy and data protection of individuals due to excessive processing of personal data by the AI system. • Arbitrary arrest or detention of individuals due to a lack of essential guarantees surrounding the AI system's use (inappropriateness, injustice, lack of predictability and due process of law). • The inadmissibility of data collected by the AI system as evidence in criminal proceedings. • A criminal or civil case brought against a law enforcement officer						
1. What is the likelihood that activities related to the AI system will lead to non-compliance with applicable laws and regulations?*		How serious would the negative impact of such non-compliance be on individuals and communities? *				

** This question focuses on the threat to individuals and communities posed by a possible lack of compliance, and should not in any way be interpreted as overriding the core principle of lawfulness. Law enforcement agencies and personnel should, of course, not perform any activities that breach applicable laws and regulations.*

Answering this question will require consultation with internal or external legal experts who are able to conduct a careful analysis of the legal and regulatory compliance risks. Before answering this question, it is important that agencies carry out a human rights impact assessment or an equivalent process in which the potential impact on human rights in relation to the AI system is identified, assessed and addressed.

MINIMIZATION OF HARM

Law enforcement agencies should prevent, eliminate, or mitigate the risk of harm to individuals and communities that can arise in the context of AI development, procurement, and use.

	LIKELIHOOD (A) 1 – Very Unlikely; 2 – Unlikely; 3 – Possible; 4 – Likely; 5 – Very likely		IMPACT (B) 1 – Insignificant; 2 – Limited; 3 – Moderate; 4 – Severe; 5 - Catastrophic		RISK LEVEL (AxB)	RISK OWNER	OBSERVATIONS
EXAMPLES: Vulnerabilities in the AI system's robustness could, for example, lead to: • Incorrect decisions due to compromised data integrity. • Suboptimal investigations due to the AI system's downtime after a cyber-attack. • Users lacking trust in the AI system's security due to fear of hacking and unauthorized disclosure of sensitive information about criminal offences or convictions. • Ineffective or delayed delivery of services to the public by the law enforcement agency (such as issuing driver's licenses or criminal record certificates) due to the AI system's unavailability.							
2. How likely is it that vulnerabilities in the AI system's robustness will result in harm to individuals or communities?		How serious would such harm be?					

EXAMPLES: Vulnerabilities in the AI system's safety could, for example, lead to:

- Death, severe illness permanent impairment, exacerbation of ailments or other physical injuries.
- Damage to or loss of equipment or property.
- Excessive or inappropriate insight by law enforcement officers into the relations, activities, preferences and behaviour of individuals or communities due to unauthorized access to data.
- The unencrypted transfer of confidential data relating to criminal convictions and offences to third parties outside the national jurisdiction.

3. How likely is it that vulnerabilities in the AI system's safety will result in harm to individuals or communities?		How serious would such harm be?				

EXAMPLES: Inaccuracies in the AI system could, for example, lead to:

- Wrongful arrests or unjustified differential treatment of individuals because the AI system generated incorrect outputs.
- Discrimination of individuals who are misrepresented or underrepresented in the AI system's training data.
- A waste of the law enforcement agency's resources due to the identification and tracking of irrelevant individuals or groups.
- A decrease in trust in the law enforcement agency due to its use of an inaccurate AI system becoming publicly known.

4. How likely is it that inaccuracies in the AI system will result in harm to individuals or communities?		How serious would such harm be?				

EXAMPLES: Using or misusing the AI system in a way that does **not preserve the well-being of people** could, for example, lead to:

- Prolonged physical or psychological suffering or other psychological harm.
- The loss or disruption of family relations.
- Loss of jobs or reduction of income.
- The emergence or exacerbation of negative perceptions among the public of law enforcement and AI due to a perceived decrease in societal well-being associated with the introduction of the AI system.

5. How likely is it that the use or misuse of the AI system will adversely affect the physical or psychological well-being of individuals or communities?		How serious would such adverse effects be?				

EXAMPLES: Using or misusing the AI system in a way that does **not preserve environmental welfare** could, for example, lead to:

- Increased energy use.
- Higher CO2 emissions.
- Increased consumption of natural resources, such as water for cooling computers and data centres.
- Penalties applied to the law enforcement agency due to the infringement of environmental legislation.

6. How likely is it that the use or misuse of the AI system will adversely affect environmental welfare?		How serious would such adverse effects be?				

EXAMPLES: Using or misusing the AI system in a way that creates excessive costs could, for example, lead to:

- The costs of using the AI system outweighing the derived benefits.
- Negative Return on Investment (ROI) or not meeting the ROI targets.
- Inappropriate use of public funds.
- Additional costs for the law enforcement agency to rectify the unnecessary increase in complexity of internal processes

7. How likely is it that excessive costs of implementing the AI system in terms of time, money, human effort and environmental impact will result in harm to individuals or communities?		How serious would such harm be?				

HUMAN AUTONOMY

Respecting human autonomy means that law enforcement agencies engage with AI in a way that safeguards humans' capacity and right to self-governance.

	LIKELIHOOD (A) 1 – Very Unlikely; 2 – Unlikely; 3 – Possible; 4 – Likely; 5 – Very likely		IMPACT (B) 1 – Insignificant; 2 – Limited; 3 – Moderate; 4 – Severe; 5 - Catastrophic		RISK LEVEL (AxB)	RISK OWNER	OBSERVATIONS
EXAMPLES: **Insufficient or inadequate human control and oversight of the AI system** could, for example, lead to: • A loss of autonomy of the individuals who interact with or are affected by the AI system's use. • Systematic human rights violations, such as structural racial discrimination or excessive surveillance. • The breakdown of democratic structures due to, for instance, a miscarriage of justice and erosion of the presumption of innocence. • The erosion of public trust in the law enforcement agency and AI.							
8. How likely is it that limitations in the ability of humans to exercise control and oversight of the AI system will result in harm to individuals or communities?		How serious would such harm be?					
EXAMPLES: Using an AI system that **limits the ability of users to make decisions independently** could, for example, lead to: • Algorithmic bias transpiring into the decisions of law enforcement officers. • The manipulation of users through features in the AI system that harness subconscious processes. • Users disregarding human input and over relying on the results of the AI system without questioning them (automation bias). • The law enforcement workforce's excessive dependency on the AI systems.							
9. How likely is it that the use or misuse of the AI system will limit users' ability to reach decisions independently?		How serious would such limitations be?					

EXAMPLES: Interference with privacy relating to data used to train the AI system could, for example, lead to:

- Individuals who were not informed about the processing of their data for training purposes lacking access and control over their personal data.
- Undue access to sensitive information by third parties.
- Distrust, discomfort, apprehension or negative sentiment by the public regarding the AI system.
- Penalties applied to the law enforcement agency, according to national regulations, due to the infringement of privacy and data protection laws

10. How likely is it that the type and amount of data used to train the AI system will have an impact on the privacy of individuals and their right to have control over their data?		How serious would such an impact be?				

EXAMPLES: Interference with privacy relating to data processed by the AI system could, for example, lead to:

- Erosion of the capacity of users, offenders, victims or other individuals to self-govern and exercise their rights due to an unnecessary and disproportionate insight into their private lives.
- Increased vulnerability and excessive exposure of individuals and communities due to the processing of large amounts of data that are unnecessary and disproportionate to the purpose of the investigation.
- Distrust, discomfort, apprehension or negative sentiment of individuals whose data protection rights are not respected.
- Penalties applied to the law enforcement agency, according to national regulations, due to the infringement of privacy and data protection laws.

11. How likely is it that the type and amount of data collected, stored and transmitted by the AI system will have an impact on the privacy of individuals and their right to have control over their data?		How serious would such an impact be?				

EXAMPLES: Using or misusing the AI system in ways that **interfere with individuals' private sphere and their capacity to self-govern** could, for example, lead to:

- Excessive insight into the political affiliations, interpersonal relations, movements and whereabouts of individuals and communities.
- The collection and analysis of information irrelevant to an investigation.
- The use of the AI system beyond its original purpose, resulting in wrongful assumptions and increased errors in decision-making.
- Distrust of the public in law enforcement and AI due to perceived excessive surveillance and observation associated with the introduction of the AI system.

12. How likely is it that the use or misuse of the AI system will interfere with individuals' private sphere and their capacity to self-govern?		How serious would such interference be?				

EXAMPLES: **The lack of knowledge by users about the AI system, its risks and limitations** could, for example, lead to:

- An increase in the AI system's vulnerability to cyber-attacks or other security threats.
- Discrimination of individuals due to the inability to monitor the quality of the data used to train the AI system and its accuracy.
- Exclusion of individuals from using the AI system due to a lack of the competencies and skills necessary to do so.
- Discomfort of users and other individuals interacting with the AI system due to the lack of awareness of existing system monitoring mechanisms.

13. How likely is it that a lack of knowledge of the AI system among its users will result in harm to individuals or communities?		How serious would such harm be?				

EXAMPLES: **The absence of public awareness regarding the use of the AI system by law enforcement** could, for example, lead to:

- Miscarriages of justice due to criminal suspects being unable to access information about the AI system used during the investigation.
- The inability of individuals affected by the AI system's use to challenge its outputs.
- The erosion of trust and confidence in society regarding the use of the AI system due to individuals and communities perceiving it as an unjustified interference with their rights.
- Unregulated or out-of-the-scope use of the AI system by the law enforcement agency due to a lack of accountability to the public.

14. How likely is it that an absence of public awareness that the AI system is being used will result in harm to individuals or communities?		How serious would such harm be?				
EXAMPLES: The inability of the users and/or people affected by the use of the AI system to understand its outcomes could, for example, lead to: • The failure of individuals harmed by the use of an AI system to obtain redress. • Miscarriages of justice due to criminal suspects being unable to challenge AI-derived evidence used against them. • A lack of foreseeability concerning the consequences of using the AI system's outputs. • The inadmissibility of AI-derived evidence due to the inability of law enforcement officers to ascertain and demonstrate the validity and integrity of such evidence in criminal proceedings						
15. How likely is it that the inability of users and/or people affected by the use of the AI system to understand how and why the AI system has reached a particular outcome will result in harm to individuals or communities?		How serious would such harm be?				

FAIRNESS

Law enforcement agencies should ensure, throughout their engagement with the AI system, a just and non-discriminatory treatment of individuals and groups and a contribution to a more equitable society.

LIKELIHOOD (A) 1 – Very Unlikely; 2 – Unlikely; 3 – Possible; 4 – Likely; 5 – Very likely		IMPACT (B) 1 – Insignificant; 2 – Limited; 3 – Moderate; 4 – Severe; 5 - Catastrophic		RISK LEVEL (AxB)	RISK OWNER	OBSERVATIONS
EXAMPLES: Training the AI system with data of insufficient quantity or quality could, for example, lead to**:** • Criminals remaining undiscovered due to the reduced accuracy of the AI system. • Unequal access to public services (such as issuing driver's licenses or criminal record certificates) by individuals belonging to groups under or misrepresented in training data sets. • Individuals belonging to under or misrepresented groups feeling stigmatized when interacting with the law enforcement agency.						
16. How likely is it that training the AI system using data of insufficient quantity or quality (for example, data with a lack of representation) will create or exacerbate inequalities or lead to discrimination?		How serious would such an impact on equality and non-discrimination be?				
EXAMPLES: Reflecting human biases in the design and development of the AI system could, for example, lead to: • Wrongful prosecution, excessive use of force and unjustified punishment of individuals belonging to communities most vulnerable to discrimination. • Individuals belonging to groups underrepresented in the development of the AI system feeling stigmatized or excluded by the law enforcement agency. • The perpetuation of harmful, false, or derogatory stereotypes.						
17. How likely is it that human biases reflected in the design and development of the AI system will have a disproportionate impact on certain individuals or groups?		How serious would such an impact be?				

EXAMPLES: Using or misusing the AI system in ways that create or exacerbate discrimination could, for example, lead to:

- Wrongful prosecution, excessive use of force and unjustified punishment of specific individuals.
- Growing tensions with individuals and groups most vulnerable to discrimination.
- Social unrest due to a lack of trust in the law enforcement agency.

18. How likely is it that the use or misuse of the AI system will create or exacerbate inequalities or lead to discrimination?		How serious would such an impact on equality and non-discrimination be?				

EXAMPLES: **The lack of consideration and engagement of vulnerable groups throughout the AI life cycle** could, for example, lead to:

- The exclusion of certain individuals and groups, such as indigenous peoples or migrants, from equal legal protection of their human rights.
- The perpetuation of harmful, false, or derogatory stereotypes.
- Distrust, discomfort, apprehension or negative sentiment by members of vulnerable groups regarding the AI system due to its perceived inaccuracy.

19. How likely is it that an insufficient consideration of and engagement with vulnerable groups throughout the AI life cycle will create or exacerbate the conditions that contribute to the vulnerability of such groups?		How serious would such an impact be?				

EXAMPLES: Using or misusing the AI system in a way that disproportionately impacts vulnerable groups could, for example, lead to:

- Discrimination of individuals belonging to vulnerable groups.
- Unequal access to opportunities by members of vulnerable groups, for instance, through unequal access to public services (such as issuing driver's licenses or criminal record certificates).
- Social unrest due to a lack of trust and an increase in negative perceptions of vulnerable groups towards the law enforcement agency.

20. How likely is it that the use or misuse of the AI system will have a disproportionate impact on vulnerable groups?		How serious would such an impact be?				

EXAMPLES: The insufficient consideration of different human characteristics and abilities during the AI system's design, development or deployment could, for example, lead to:

- A lack of access to law enforcement services by certain individuals due to the underperformance of the AI system with regard to their age, gender, ability, or other characteristics.
- Individuals with disabilities feeling stigmatized or excluded by the law enforcement agency when interacting with the AI system.
- A lack of confidence, irritation, anxiety or negative sentiment by AI system users whose features (for example, voice, mimics and gestures) are not recognized by the AI system.
- The inefficiency of digital services provided by the law enforcement agency with the support of the AI system.

21. How likely is it that an insufficient consideration of different human characteristics and abilities during the AI system's design, development or deployment will create or exacerbate disadvantages for certain individuals or groups?		How serious would such disadvantages be?				

EXAMPLES: The absence of measures to allow the AI system's users to challenge its outputs could, for example, lead to:

- Failures, inaccuracies and biases of the AI system remaining undetected and uncorrected.
- Annoyance, irritation or discomfort by the AI system's users for not being able to report their experience with the AI system (such as bugs, lags and inefficiencies).
- A lack of trust by the law enforcement personnel in the AI system's outputs.

22. How likely is it that a lack of technological and/or organizational measures to allow AI system's users to challenge its outputs will have a negative impact on individuals or communities?		How serious would such adverse effects be?				

EXAMPLES: The absence of measures to allow the people affected by the use of the AI system to challenge its outputs could, for example, lead to:

- Failures, inaccuracies and biases of the AI system remaining undetected and uncorrected.
- Discriminatory practices not being identified and addressed.
- The erosion of public trust in the law enforcement agency and AI due to the inability to raise perceived or alleged human rights violations.

23. How likely is it that a lack of technological and/or organizational measures to allow the people affected by the use of the AI system to challenge its outputs will have a negative impact on individuals or communities?		How serious would such adverse effects be?				

EXAMPLES: **Insufficient access to redress** by those who suffer negative effects as a result of the AI system's use could, for example, lead to:

- A breach of the right to access to justice and redress to those who claim to have suffered human rights violations as a result of the use of the AI system.
- Unequal access to effective judicial remedies by certain individuals in the absence of adequate procedures to allow groups of victims to present claims for reparation.
- A feeling of unsafety and physical and psychological distress by those who are affected by the use of the AI system or their relatives.

24. How likely is it that insufficient access to redress for those who suffer negative effects as a result of the use of the AI system will have a negative impact on individuals or communities?		How serious would such adverse effects be?				

Annex A:
Glossary of key terms[2]

RISK ASSESSMENT

A risk assessment is the process within risk management that is designed to identify and evaluate the risks to individuals and communities that may emerge during the life cycle of a specific AI system.

RISK

A risk consists of a threat to individuals and communities posed by potential circumstances and events related to the AI system that may arise at any stage of the life cycle. These are circumstances and events that have not yet occurred: they are hypothetical scenarios, to be used where there is uncertainty surrounding the effects of implementing a certain AI system for a particular use case.

In this Risk Assessment, the risk level is calculated using the following formula

Risk Level = Likelihood x Impact

This formula gives us the risk level as a function of the probability of an occurrence (the *likelihood) and the severity of the consequences of that occurrence (the impact).*

RISK OWNER

The risk owner is the entity or individual within the law enforcement agency with the authority and accountability to administer and/or respond to a risk.

MITIGATION MEASURES

Mitigation measures are any processes, policies, practices, or devices intended to decrease the risk level by: (1) decreasing the probability and/or severity of the impact of a particular event or situation; or (2) eliminating the source of the risk. It is important to acknowledge that these mitigation measures may not always succeed in reducing the risk level as intended, due to unforeseen events or insufficient understanding of the risks in question.

LIKELIHOOD

Likelihood refers to the probability of a certain circumstance or event occurring. In this Risk Assessment, likelihood is defined on a scale of 1 to 5, corresponding to the following:

1. **Very Unlikely:** There is a very low chance that the circumstance or event will occur. It would only happen under exceptional circumstances or in rare cases. This level generally corresponds to risks that, while possible, are considered negligible.

2. **Unlikely:** The circumstance or event is not expected to occur frequently or in the normal course of events. This level generally corresponds to occurrences that, while no longer considered negligible, are unusual or uncommon.

3. **Possible:** There is a fair chance the circumstance or event will occur. It may happen occasionally, and may be triggered by certain conditions, but it is not something that is expected to happen consistently or frequently.

4. **Likely:** The circumstance or event is expected to occur frequently or in the normal course of events. It may not happen every time, but there is a substantial probability that it will occur.

5. **Very likely:** The circumstance or event is almost certain to occur. It is expected to happen most of the time, barring exceptional circumstances that prevent it.

IMPACT

Impact refers to the severity of the potential harm or negative effect that the circumstance or event would have on individuals and communities if it occurred. For the purposes of this Risk Assessment, "individuals and communities" refers to any stakeholder that may be affected by the use of the AI system. |► *Learn more about how to identify the stakeholders in the* ***Principles for Responsible AI Innovation*** *and in the* ***Responsible AI Innovation in Action Workbook****.*

1. **Insignificant:** If the circumstance or event were to occur, it would have minimal or no real impact on individuals or communities. It may cause slight inconvenience or minor disruption, but it would not lead to substantial damage or harm.

2. **Limited**: If the circumstance or event were to occur, it would cause some disruption or damage, but the effects would be relatively contained and manageable. It might lead to a temporary setback or require some effort to correct, but it would not cause long-term or widespread harm.

3. **Moderate:** If the circumstance or event were to occur, it would cause a significant level of disruption or harm. This could involve substantial loss of resources or major inconveniences. However, recovery would be relatively straightforward given the right corrective action.

4. **Severe**: If the circumstance or event were to occur, it would lead to serious harm or disruption. This could involve major losses, significant harm to individuals, severe damage to society or the environment, or considerable legal or ethical implications. Recovery could be difficult, costly, or time-consuming.

5. **Catastrophic:** If the circumstance or event were to occur, it would cause devastating harm or disruption. This could involve massive losses, extreme harm to individuals, existential threats to society or the environment, or widespread violation of laws or ethical norms. Recovery may be impossible, or may require a complete overhaul of the system or the organization.

Annex B:
Risk mitigation measures

Risk mitigation measures are designed to decrease the level of risk. They consist of processes, policies, practices, or devices intended to reduce the probability and/or severity of a particular event or situation, or eliminate the source of the risk.

Annex B presents a set of examples that intend to support law enforcement agencies in identifying suitable risk mitigation strategies. The examples are divided into categories, each referring to the relevant core principles for responsible AI innovation. The list below is only illustrative and not exhaustive.

LAWFULNESS

- Oversight mechanism ensuring the AI system's compliance with legal obligations under national law and international human rights law, including compliance with the principles of rule of law, legality, necessity, and proportionality.
- Regular and systematic evaluation of the AI system during development, deployment, and use.
- Internal procedure for monitoring AI-driven decisions regarding evidence-gathering.

MINIMIZATION OF HARM

- Technical measures built into the AI system ensuring the encryption of sensitive and confidential data on storage and transmission.
- Organizational measures to regularly evaluate the effectiveness of applied encryption technologies.
- Awareness-raising among law enforcement personnel about the necessity for using encryption to avoid unauthorised access, disclosure, accidental or unlawful alteration, loss, or destruction of data during transmission and storage.
- Formal procedure for granting, monitoring, and revoking permissions to access the AI system (access management system)

- Regular and systematic logging of user access to the AI system.
- Regular and systematic logging of user activity in the AI system, such as inputs and changes.
- Regular and systematic control of anomalies' logs.
- Formal control procedure for testing the AI system's inputs and outputs.
- Organizational measures to assign responsibility to a specific team for measuring and evaluating the impact of the AI system on individuals, communities, and the environment.
- Consultations with public and private sector stakeholders and civil society organizations on perceived negative impacts of the AI system.

HUMAN AUTONOMY

- Formal review procedures within the data model management system.
- Formal procedures for monitoring and controlling the accuracy of the AI system's inputs and outputs.
- Formal procedures for human intervention, for example, through governance mechanisms such as human-in-the-loop (HTL), human-on-the-loop (HOTL) and human-in-command (HIC).
- Log trail showing the identity of the authorizing decision-maker.
- Following Privacy-by-Design and Security-by-Design principles when designing and developing the AI system.
- Checklist for inclusion of privacy and security measures in the AI system for designers and developers.
- Formal procedures for testing and evaluating the explainability and interpretability of the AI system throughout the life cycle.
- Regular audit of the AI system by an independent external auditor.

- Mandatory training programmes for all users of the AI system in basic concepts related to AI, ethical policing, human rights, and the principles for responsible AI innovation.
- Mandatory training programmes specialized for different units and functions within the law enforcement agency on how to use the AI system.
- Review of training programmes and testing procedures based on potential incidents related to the use of the AI system.
- Training of relevant law enforcement personnel in handling feedback and complaints regarding the AI system.
- Public disclosure of the AI systems used by the law enforcement agency, for example, through AI registries.

FAIRNESS

- Formal procedure to review the quality and quantity of training data sets and detect biases in the AI system.
- Organizational measures to assign responsibility to law enforcement personnel with specific functions to monitor the AI system for bias throughout the AI life cycle.
- Formal procedures or processes for regular and systematic bias monitoring regarding the AI system's outputs.
- Formal procedure to include stakeholders in testing the AI system for bias in all AI life cycle phases.
- Complaints mechanism for users and individuals affected by the use of the AI system to claim redress in case of perceived harm or adverse impact.
- Dissemination of information to the AI system's users on how to provide feedback on the AI system.
- Dissemination of information to users and individuals affected by the use of the AI system on how to submit a complaint regarding the use of the AI system.

ENDNOTES

1 National Institute of Standards and Technology. (2012). Guide for Conducting Risk Assessments (NIST Special Publication (SP) 800-30 Rev. 1). Accessible at https://doi.org/10.6028/NIST.SP.800-30r1

2 International Organization for Standardization. (2009). *ISO Guide 73, Risk management—Vocabulary. Accessible at* https://www.iso.org/obp/ui/#iso:std:iso:guide:73:ed-1:v1:en

OTHER SOURCES

International Organization for Standardization. (2018). ISO 31000:2018 - Risk management - A practical guide. Accessible at ISO - ISO 31000:2018 - Risk management — A practical guide

National Institute of Standards and Technology. (2012). Guide for Conducting Risk Managements (NIST Special Publication (SP) 800-30 Rev. 1). Accessible at SP 800-30 Rev. 1, Guide for Conducting Risk Assessments | CSRC (nist.gov)

Report of the Special Representative of the SecretaryGeneral on the issue of human rights and transnational corporations and other business enterprises, John Ruggie. (2011). Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. (A/HRC/17/3) Accessible at https://documents.un.org/doc/undoc/gen/g11/121/90/pdf/g1112190.pdf?token=ZxWmrxeESbzRgX1b9h&fe=true

United Nations Human Rights Office of the High Commissioner (2023). Taxonomy of Human Rights Risks Connected to Generative AI. Accessible at https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf

How to cite this publication: UNICRI and INTERPOL. (Revised February 2024). Toolkit for Responsible AI Innovation in Law Enforcement: **Risk Assessment Questionnaire.**

2025 Update: This concerns a correction to minor errors contained in the previous version.

www.interpol.int
www.unicri.it

INTERPOL_HQ

@INTERPOL_HQ
@UNICRI

INTERPOL HQ
UNICRI

INTERPOL
UNICRI

@INTERPOL
@UNICRIHQ

www.ai-lawenforcement.org