



**RESPONSIBLE
AI INNOVATION IN
LAW ENFORCEMENT**
AI Toolkit

Principles for Responsible AI Innovation



Funded by
the European Union

REVISED FEBRUARY 2024

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, nor do they imply any endorsement by, INTERPOL or the United Nations, their member countries or member states, their governing bodies, or contributory organizations. 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

This document lists and explains the principles for responsible AI innovation. These principles are the foundation for the entire AI Toolkit, and they are designed to guide law enforcement agencies across the world in integrating AI systems into their work in ways that align with good policing practices and AI ethics, and respect human rights. This document also briefly explains how to put them into practice. For a more in-depth explanation, see the Responsible AI Innovation in Action Workbook.

WHEN

The *Principles for Responsible AI Innovation* are designed to be followed throughout the AI life cycle. It is recommended that agencies gain a thorough understanding of these principles from the beginning of their involvement with an AI system, and that they refresh or expand their knowledge of each of the principles throughout the process.

WHO

The principles are relevant to all stakeholders in a law enforcement agency, within their functions and capacities.

Table of Contents

DISCLAIMER	1
OVERVIEW	2
Understanding the principles	4
The principles for responsible AI innovation	6
1. LAWFULNESS	6
2. MINIMIZATION OF HARM	10
3. HUMAN AUTONOMY	15
4. FAIRNESS	22
Putting the principles into practice	30
UNDERSTANDING AND APPLYING THE PRINCIPLES	31
IDENTIFYING AND ENGAGING WITH THE RELEVANT STAKEHOLDERS	32
CHECKING THE RESULTS	33
REPEAT (IF NEEDED)	33
<i>Annex: Want to learn more?</i>	34
<i>Endnotes</i>	39

Understanding the principles

Modern-day policing rests on a bedrock of principles that are central to sustaining effective and fair criminal justice systems.¹ Many of the principles for responsible AI Innovation included in this document will therefore undoubtedly be familiar to the law enforcement community, as they are ingrained in the national laws and international standards that are at the foundation of all modern police work. The purpose of this document is thus not to reinvent the wheel, but to use established principles to provide law enforcement agencies with a framework for how to think about AI, and to explain how responsible AI innovation, through a principled approach, can be successfully implemented in a law enforcement context.

Due to the importance of law enforcement, the crucial role it plays in society, and the impact it has on individuals' lives, agencies and officers have a duty to follow the highest standards of conduct in the exercise of their functions.² These high standards should also apply to law enforcement agencies that are currently developing, procuring or using AI systems or seeking to integrate AI systems into their work in the future. The following **five core principles for responsible AI innovation provide the law enforcement community with a foundation for a principled approach to AI: 1) Lawfulness; 2) Minimization of Harm; 3) Human Autonomy; 4) Fairness; 5) Good Governance.**



Image by utah51 - stock.adobe.com

These core principles define responsible AI innovation in law enforcement. In other words, **responsible AI innovation in law enforcement consists of developing, procuring, and deploying AI systems in a way that is lawful, minimizes harm, respects human autonomy, is fair, and is supported by good governance.**

To realize the five core principles, law enforcement agencies can rely on a set of instrumental principles that complement each of the core principles, as shown in the figure below:

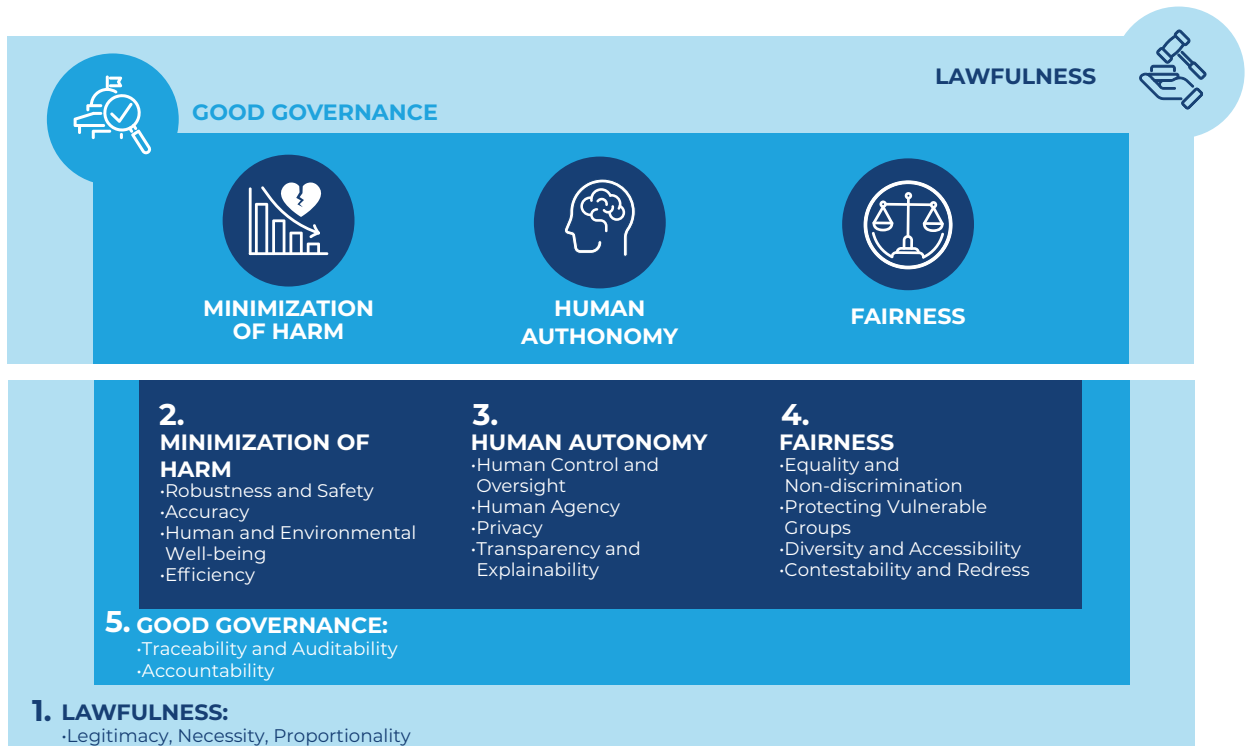


Figure 1 - Core and Instrumental Principles

WANT TO LEARN MORE?

See the [“The foundations of the core and instrumental principles”](#) section in the annex.

The instrumental principles help to achieve the core principles.³ Sometimes instrumental principles will be conflicting, or it will not be possible to fulfil them entirely. In such cases, other instrumental principles can be used to ensure that each of the core principles remains protected. This will be explained in more detail later in the document.

As shown in the figure above, it is important to note that lawfulness underpins all the other core and instrumental principles in this context. This is because the law, including human rights law, provides both the foundation and the limit for law enforcement actions, including those involving AI.

While some of the instrumental principles address specific human rights issues such as privacy and non-discrimination, agencies should still be mindful of the broader impact their AI activities may have on human rights.

The principles for responsible AI innovation

The principles – both core and instrumental – designed to guide law enforcement agencies toward responsible AI innovation are explained in the following subsections. The principles are described in such a manner that allows them to be adapted to the diverse contexts in which law enforcement operates.

1. LAWFULNESS

Like any other activity that law enforcement agencies carry out as part of their mission to prevent, detect, and investigate crime, their engagement with AI systems needs to be lawful.⁴ This means that **agencies must follow the applicable laws and regulations throughout the design, development, and use of AI systems.**

Lawfulness entails respecting the specific laws and regulations that apply in the territory where law enforcement agencies operate. These will vary across regions, countries or districts and will change over time, especially since AI is a rapidly evolving field.⁵

Respecting human rights is also an essential part of lawfulness. Law enforcement agencies have a general obligation to safeguard human dignity and uphold the human rights of all persons.⁶ Therefore, responsible AI innovation in law enforcement requires agencies to determine and avoid or mitigate the impact that developing, procuring or deploying an AI system may have on the rights of any individual – whether a victim of crime, a suspect or criminal, law enforcement personnel or a member of the general population. This includes the rights recognized and established in international law, which comprise basic and adaptable standards for human rights protection that benefit from global consensus,⁷ as well as those specified in regional and national laws. To help determine if an AI system affects human rights, law enforcement agencies should involve ethics and human rights experts in carrying out a human rights impact assessment. ▶ Learn more in the Organizational Roadmap and in the Risk Assessment Questionnaire.

The following principles are instrumental to lawfulness:

- [Legitimacy, necessity, and proportionality](#)

LEGITIMACY, NECESSITY AND PROPORTIONALITY

Developing, procuring or deploying an AI system in law enforcement may have an impact on the rights of citizens, law enforcement personnel, victims and suspects of crime, criminals, or other individuals. For lawful AI innovation, law enforcement agencies should ensure legitimacy, necessity, and proportionality whenever they engage with AI systems in ways that could have an impact on human rights.

Legitimacy means that law enforcement agencies should only interfere with people's rights when they have a valid reason to do so, based on domestic law and in line with international standards.⁸ This means that, for any interference with human rights, law enforcement agencies need to fulfil two requirements from the beginning:

- having a *legal basis for that interference*; and
- *following a legitimate goal* such as safeguarding the life and safety of individuals and society.

Necessity means that law enforcement agencies should only interfere with people's rights when such interferences are needed to fulfil the identified legitimate goal. This means that, even when the pursued goal is legitimate, agencies should ensure that it cannot be achieved without interfering with human rights. They should also note that while interference may be necessary at first, it may become unnecessary if the goal is achieved or can no longer be achieved in a lawful way.

Proportionality means that law enforcement agencies need to balance the interference with human rights against the reason for doing it (the legitimate goal). This implies that interferences must always correspond to the least intrusive way of achieving such a goal⁹ and that the negative effects they have on people's rights must be balanced against the legitimate goal pursued.¹⁰ This balancing exercise is also closely connected with the core principle of [fairness](#).

Law enforcement agencies are very familiar with these principles and they are already part of their daily practices, guiding them, for example, when choosing an investigatory measure or using force against people.¹¹ They are general legal requirements for limitations to human rights, although they may appear in different legal or operational frameworks under separate names, such as "strict necessity" or "adequacy". ▶ *Learn more about limitations to human rights in the **Introduction to Responsible AI Innovation**.*

COMMON QUESTION**When should legitimacy, necessity and proportionality be checked?**

Legitimacy, necessity, and proportionality should be evaluated before law enforcement agencies decide to perform an activity that could potentially interfere with human rights – for example, before they decide to implement an AI system to scan files from the devices of suspects apprehended during investigations. However, this should not be a one-off exercise. Law enforcement agencies should ensure that all activities in the AI system’s life cycle fulfil the requirements for legitimacy, necessity, and proportionality, and should regularly reassess the situation. For example, if the evidence for the crime has been collected and a suspect has been indicted, the confiscated devices should not be scanned for other unrelated information.

Because legitimacy, proportionality and necessity are interconnected, it is helpful to adopt a step-by-step approach to ensure that these principles are being properly addressed. The example below explains this approach with a hypothetical scenario.

PRACTICAL EXAMPLE**A step-by-step assessment of legitimacy, necessity, and proportionality**

Suppose a law enforcement unit wants to procure a text processing AI system that will allow investigators to scan devices that are confiscated during child sexual abuse investigations. This AI system will be able to flag any conversations that may contain evidence of crimes against children. Before procuring this system, it is imperative to ensure that it will be used in a lawful manner by the investigators and that agency-wide guidance is available regarding the rules to be applied while using AI systems. Ideally, the agency would have support from internal or external legal experts for these situations. |► *Learn more about the people and expertise required for responsible AI innovation in the **Organizational Roadmap**.*

The following step-by-step approach illustrates how this may be undertaken.

Step 1: Identify possible interferences with human rights.

Before procuring the AI system, the agency should conduct a human rights impact assessment. Implementing this AI system will probably interfere with the privacy of different people – not only the suspect of the crime but also anyone the suspect exchanged messages with, including the victims of child sexual abuse. While there are other potential human rights interferences at stake, such as the right to equality, we will focus on privacy for this example.

Step 2: Identify the legal basis for possible interferences.

As using the AI system may interfere with the right to privacy, it is important to ensure that the law provides a legal basis for this. If the law allows interference with the right to privacy, it is likely that the legal framework includes certain requirements for this interference. Lawfulness also means that the agency and the officers must adhere to these requirements. It is recommended

that agencies, through consultation with their legal experts, provide some guidance to officers in charge of the investigations regarding the rules that they need to follow in an operational setting. If the applicable legal framework does not provide a clear basis for deploying the AI system, it should not be procured or deployed.

Step 3: Establish whether there is a legitimate goal for the interferences.

Apart from determining whether there is a general legal basis for procuring or deploying the AI system, it is important to consider the goal that the agency aims to achieve by using this system which may lead to interferences with human rights. Lawfulness requires that such a goal is legitimate according to the law. Let us suppose that the unit aims to improve the effectiveness and the speed of investigations into child sexual abuse to safeguard more children and apprehend more perpetrators. This would be a legitimate goal as it aligns in principle not only with national but also international standards for ensuring public safety and protecting the life and safety of children. However, if the agency aims to use this AI system for speculative purposes or purely to test it in the field, the objective would not be legitimate.

Step 4: Assess the necessity of the interferences.

If the activity aims to achieve a legitimate law enforcement objective foreseen in the law, the next step is determining whether the interferences are necessary. This can be done by assessing whether the use of the AI system is at all necessary for the legitimate goal of conducting more effective criminal investigations and thereby protecting public order and the life and safety of individuals. The measures taken by law enforcement agencies must be necessary for fulfilling that specific purpose.

Another important component of this concept is reassessing the necessity while the AI system is in use. For example, if the investigation is closed, officers should stop using the AI system to scan the apprehended devices.

Step 5: Assess the proportionality of the interferences.

After establishing the necessity of using the AI system, the agency must ensure that the use of the AI system is proportionate to the goal of protecting the life and safety of individuals and fighting crime effectively, i.e., *the legitimate goal*.

To this end, the agency should ensure that there are no alternative measures with a lower impact on the right to privacy that could be taken to improve the efficiency of investigations. For instance, if there are other AI systems available on the market that offer better privacy protection, it would be advisable to opt for one of these.

The agency also needs to strike a balance between the aim pursued and the use of an AI system as well as its potential consequences in a specific case.¹² Any negative impact on the right to privacy cannot be worse than the reason for the interference. In most cases, this requires weighing against each other the very real effects of the planned use of the AI system on the right to privacy and the legitimate goal. For example, if the AI system has automatically collected the names or other personal information related to the suspects, their network, and the children that were potentially victims of abuse, and stored it in a non-secured way, then the negative consequences of using the AI system on the life and safety of individuals may be higher than the negative consequences of not using it.

2. MINIMIZATION OF HARM

Minimizing harm is a fundamental goal of policing. The essence of law enforcement is to protect people and society against illegal acts, including by preventing and combatting crime.¹³ The same principle is crucial in the context of responsible AI innovation. In this context, **minimization of harm means that law enforcement agencies prevent, eliminate, or mitigate the risk of harm to individuals and communities that can arise in the context of AI development, procurement, and use.**

To do so, the first step is to define and identify the possible harm to individuals, society and the environment that may result from procuring, developing, and deploying an AI system. Assessing the human rights impact of an AI-related activity, as described under lawfulness, can in part serve this purpose. However, “harm” is a broader concept than human rights interference. It covers all the adverse consequences of an action or a policy on the physical, mental, social, or economic well-being of people, society, and the environment, even if they do not amount to an interference with individual rights.

Once the potential for any sort of harm is identified or actual harm is detected, law enforcement agencies should adjust their action so that the harm is avoided or stopped, or at least mitigated. In calibrating the minimization of harm, agencies also need to consider the consequences of alternative actions or policies: in other words, the risks, and benefits of the alternatives and who would be affected should be evaluated.

As actions involving AI have the potential to cause harm, and harm can sometimes be justified, the principle is formulated as minimization of harm rather than “do no harm”. However, when the harm at stake corresponds to a human rights interference, the principles of legitimacy, necessity and proportionality should be applied to determine the appropriate course of action.

The following principles are instrumental to minimization of harm:

- [Robustness and Safety](#)
- [Accuracy](#)
- [Human and environmental well-being](#)
- [Efficiency](#)

ROBUSTNESS AND SAFETY

Robustness and safety imply that AI systems can maintain consistency across different contexts and identify and prevent potential risks of harm, and that they are protected against attacks and overall do not pose a threat to the physical or mental well-being of individuals, their property, or the environment. Given the central role that robustness and safety play in preventing and minimizing the risks of harm posed by the use of AI systems, responsible AI innovation in law enforcement requires agencies to verify that the AI systems they are developing and using are built in line with these principles.

More specifically, **to ensure robustness, law enforcement agencies should confirm that the AI systems they intend to use are both reliable and secure.**¹⁴

- The *reliability* of an AI system is its ability to perform its intended function adequately and consistently over time, with different inputs and in different contexts. A reliable AI system is one that is capable of coping with changes in its environment while still maintaining a consistent performance.
- The *security* of an AI system relates to how protected it is against potential attacks. A secure AI system is one that maintains its integrity and the confidentiality of any data in case of attempts at exploitation by adversaries.

The types of harm that can derive from a lack of robustness include modifications to data, unauthorized access to software, hardware and infrastructure, or changes in the behaviour of the AI system resulting in erroneous decisions or causing the system to shut down.

Law enforcement agencies also need to make sure that any AI systems they aim to use are safe, meaning that they include sufficient safeguards to prevent unacceptable harm and minimize unintentional and unexpected harm. Safety, therefore, relates to all risks of harm posed by AI systems, including the risks that stem from dual use of the system or any risks that arise when the system encounters a problem or fails. Ultimately, the definition of safety is ensuring that the system does not put individuals, goods, or the environment in danger.¹⁵

The principles of robustness and safety can be translated into technical and organizational measures that should be put in place, and the effectiveness of these measures needs to be regularly checked throughout the AI life cycle. These principles are therefore fundamental for law enforcement agencies regardless of the way they engage with a specific AI system. In other words, they are relevant whether developing a system or procuring it from a third party, and while the system is in use.

ACCURACY

Accuracy corresponds to the degree to which an AI system can make correct predictions, recommendations, or decisions.¹⁶ It is important that agencies verify that any system they are developing and/or intend to use is highly accurate, as using inaccurate AI systems can result in various types of harm. For example, if an AI system used for crime detection has a low accuracy rate, it could potentially cause law enforcement officers to be misled into responding to a location where no actual crime has occurred. This could be detrimental to both law enforcement agencies and society as a whole, as it would result in the unnecessary waste of valuable and often scarce resources. Therefore, before deploying an AI system into mainstream application in the law enforcement context, such system needs to be subject to rigorous and scientific testing.

The accuracy of an AI system is dependent on the way the system was developed, and in particular the data that was used to train it. In fact, training the system with sufficient and good quality data is paramount to building a good AI model. In this regard, agencies should be particularly mindful of the origin and composition of the training data, both when procuring an AI system or developing it internally. In most cases, it is preferable that the training data relates to the same or a similar context as the one where the AI system will be used. ▶ *Learn more about Model Performance and Data Requirements in the **Technical Reference Book**.*

Accuracy can also vary according to the context in which the system is used and the input it receives. This is why organizational measures, such as requiring testing by independent third parties before buying an AI system and monitoring the system's accuracy throughout its life cycle, are important. Moreover, it is recommended that law enforcement personnel are trained to interpret and question the system's outputs. ▶ *Learn more about the recommendations on People and Expertise for responsible AI innovation in law enforcement in the **Organizational Roadmap**.*

PRACTICAL EXAMPLE

Understanding accuracy with facial recognition technology

Facial recognition technology is widely used in law enforcement to support the identification of subjects of interest by matching an unknown face to one whose identity is known. It supports the identification of individuals on a one-to-one verification basis – matching a face to an ID card, for example – or a one-to-many basis – comparing an unknown face to a database of known faces to search for their identity.

Law enforcement agencies must carefully consider the development and use of post-event facial recognition technology, given its widespread adoption across many countries and its potential benefits for crime prevention and investigation, as well as the concerns, controversy, and negative consequences it has generated. It is important to ensure the accuracy of such systems and, more broadly, the accuracy of the identification process as whole.¹⁷ To this end, it is crucial that agencies understand the factors that may influence the system's ability to make correct potential matches, some of which are set out below:

Training data sets

Accuracy can be affected by the quality and quantity of the data that was used to train the facial recognition system. To develop a good algorithm, it is not enough to have a big data set – the training data set also needs to represent the population in which it will be deployed. A facial recognition model trained with an unrepresentative data set will have fewer examples to learn from for certain categories of the population compared to others. For example, if the data set used for training does not have enough images of faces of people from a certain racial or ethnic minority, the AI system will learn less nuances for those particular categories which may result in lower accuracy when identifying people from those groups.

Threshold adjustments in the model during development

A facial recognition system works by calculating the probability of two faces belonging to the same person. This probability is then converted into a classification label – either “match” or “no match” – depending on the classification threshold. This threshold is defined by developers to produce the greatest number of correct matches and no matches – i.e., to optimize accuracy. Increasing the decision threshold, for example to 99%, means that the system classifies two pictures as a match when it is 99% confident that they belong to the same person. This reduces the number of incorrect matches (false positives) but increases the possibility of missing an actual match (false negatives). Lowering the threshold will make the model consider that two pictures match more often, therefore decreasing false negatives. However, this also leads to a larger number of incorrect matches (false positives). If the model is trained with an unrepresentative data set, the number of incorrect matches will be higher for under-represented categories than for categories better represented in the data.¹⁸

Conditions of usage

The accuracy of facial recognition systems also varies widely depending on the quality of the image that is fed into the system for analysis. In ideal conditions (in terms of lighting, positioning, and image resolution), certain facial recognition systems can achieve accuracy scores above 99%. However, their accuracy rate can drop to below 80% if fed with low-quality images, such as side-view images or images captured with low-quality webcams or ATM-style registered traveller kiosks.¹⁹

In terms of responsible AI innovation, the examples above show that:

Firstly, during development, deficiencies in the training data may affect an AI system’s ability to accurately identify people from certain groups, and that human decisions play an essential role in determining to what extent this is the case. When an AI system is developed to analyse data that relates to people, this can create a disproportionate and unfair burden on individuals that belong to certain groups. This means that the accuracy of the AI system can also affect the principle of fairness.

Secondly, accuracy may vary according to the context in which the AI system is used. For that reason, law enforcement agencies should be mindful of the conditions in which a certain AI system is intended to be used so they can properly understand the risks and benefits that using the system may bring. For example, using low-quality images collected from public spaces for real-time facial recognition may not generate a good investigative lead and could negatively impact public perception of the use of this technology in law enforcement.

HUMAN AND ENVIRONMENTAL WELL-BEING

The principle of human and environmental well-being entails law enforcement agencies preserving and improving the welfare of people and the environment in their AI innovation journey.

This consideration is partially ensured by the principles of [robustness](#), safety, and [accuracy](#). However, human and environmental well-being is a broader principle, as it implies that agencies should examine the full spectrum of direct and indirect consequences of their AI-related activities and aim for the improvement of well-being. By combining societal and environmental sustainability issues, this principle can facilitate discussion and consideration of matters such as energy consumption and the use of resources during all phases of the AI system's life cycle. In this sense, it is also connected with the principle of [efficiency](#).

PRACTICAL EXAMPLE

Using image processing systems in public spaces

Let us imagine that a law enforcement agency wants to use an object recognition system that uses images collected by CCTV cameras in public places, to detect security threats based on certain movements of people and objects.

Introducing an AI system like this in a responsible way requires the agency (among other considerations) to account for **how it will impact people's well-being** – for example, how the inhabitants of the area in question will feel about it. Certain inhabitants may feel safer if such systems exist, whereas others may experience discomfort and a feeling of being “watched”. Societal well-being may also be affected, depending on how people respond to having fewer law enforcement officers on the streets than if the AI system was not in place.

Understanding this will help the law enforcement agency decide if and how the AI system should be implemented. If the introduction of such a system result in a significant decrease in societal well-being, there is a possibility that negative perceptions among the public of law enforcement and AI could emerge or be reinforced. This would decrease trust in law enforcement and therefore compromise officers' work. However, this can change over time: public perceptions and attitudes about the use of such image processing systems could improve if they are adequately informed about how they work and trust that their rights will be safeguarded throughout the process.

Another aspect of the principle of well-being is the capability of the AI system to **function in the most environmentally friendly way possible**. An object recognition system with several cameras in different parts of a city recording 24/7 results in large volumes of footage, especially if it works with high quality images. The energy costs of an AI system will be influenced by factors such as the volume of data collected, the method of transfer and the location and duration of storage. For that reason, ensuring environmental well-being requires developing systems that collect the least amount of data possible and store data for the shortest period possible in line with existing national and regional data protection laws.

EFFICIENCY

Efficiency in AI innovation means that law enforcement agencies make sure that there is a favourable ratio between the costs and the benefits of using a certain AI system in terms of time, money, human effort, and the impact on the environment.

One of AI's biggest promises is efficiency. Using AI systems can allow complex tasks to be completed in a faster, easier, and less-resource intensive manner. However, costs are incurred at all stages of the AI system's life cycle. For example, agencies need to spend money, time, and human and environmental resources on developing, procuring, and deploying a good system, including training personnel to use and monitor it, and purchasing adequate hardware for it to run. The efficiency principle requires agencies to determine whether the benefits of using the system outweigh the costs.

This is particularly relevant because, especially in the current era of digital transformation, agencies may feel compelled to adopt AI systems even when the benefit is unclear or when it adds an extra layer of unnecessary complexity to an existing internal process. If a process becomes unnecessarily complex, it can result not only in more errors but also in additional spending to rectify the negative consequences arising from these errors. Conducting an agency-wide needs and capabilities assessment before deciding whether to integrate AI systems into the current structure is thus an important process to enable responsible AI innovation. [▶ Learn more about the recommendations on Processes for responsible AI innovation in law enforcement in the Organizational Roadmap.](#)

3. 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, whether the law enforcement personnel using the tool, victims of crime, suspects, criminals, or the public in general.

This principle is rooted in the idea that every human has an inviolable value simply by virtue of belonging to a species capable of rationality. It is the basis of globally recognized and valued concepts such as human dignity and human rights. Safeguarding human autonomy entails protecting the independence and dignity of every individual or group that interacts with or is affected by the use of an AI system.

The following principles are instrumental to human autonomy:

- [Human control and oversight](#)
- [Human agency](#)
- [Privacy](#)
- [Transparency and Explainability](#)

HUMAN CONTROL AND OVERSIGHT

In the context of AI, human control and oversight are the ability and opportunity for humans to adequately supervise, engage and interfere with an AI system during its development and use. To ensure human control and oversight, **law enforcement agencies are advised to verify that the AI systems they currently use or intend to use are built with the functionalities needed to ensure that humans remain in charge during use, as well as to confirm that the necessary organizational structures are in place to ensure that humans have the last word regarding certain decisions.**

WANT TO LEARN MORE?

See the "[Human-in-the-loop, human-on-the-loop, human-in-command](#)" section in the annex.

The terms human-in-the-loop, human-on-the-loop and human-in-command are often used to refer to the governance mechanisms needed to set up these functionalities and structures.²⁰ However, to adequately safeguard human autonomy in decision-making and lawfulness, it is essential that the humans "in-the-loop", "on-the-loop" or "in-command" have a proper understanding of the AI system they are interacting with. This relates to the principles of transparency and explainability. It is equally important that humans and the structures that they are part of are independent, which relates to the principle of accountability.²¹

Upholding human control and oversight of AI systems is particularly important in the law enforcement context. This is especially true considering that the work of law enforcement agencies is at the very core of the functioning of society, justice, and political systems, and therefore has a significant influence on individuals and their rights. For that reason, AI systems with a high degree of autonomy – meaning, those which are able to make decisions about the "real world" and act on them without human supervision and intervention – are generally not recommended, as their decisions can have a direct impact on people's lives. Ensuring that these principles are upheld is particularly important for the personnel interacting with AI systems, as they are ultimately responsible for any decisions taken with the assistance of AI.

HUMAN AGENCY

Human agency is the ability of a person to act upon their own decisions and pursue their goals without manipulation or force. To protect human agency in the context of AI innovation, **law enforcement agencies need to ensure that the AI systems they aim to use do not compromise the ability of the users of those systems (law enforcement officers, other personnel, citizens, etc.) to act and make decisions independently.**

Human agency can be challenged if individuals or institutions are over reliant on AI systems, disregarding human input when it may be relevant or even necessary. For example, if an AI system is used in a certain process, agencies should in most cases ensure that the system is genuinely supporting or improving the decisions taken by the officers in charge of the process, instead of making those decisions for them. This also entails training the officers, so they know how to engage properly with the AI system.

At the institutional level, an over-reliance of law enforcement agencies on AI systems has the potential to disrupt the balance between AI innovation and the associated risks by diverting resources away from other policing methods that do not carry similar risks. Therefore, law enforcement agencies should ensure that their workforce does not become overly dependent on AI systems. For example, recruits should develop and maintain well-rounded policing capabilities which are not dependent on AI systems, even if, at some point in the future, the use of AI systems for a specific activity is the standard practice.

Human agency can also be affected if the AI system is used to limit people's access to information or opportunities, or if it is deployed to manipulate and/or control individual behaviour. An AI system that interacts with the public – for example, an AI chatbot used to help people submit a question or complaint to the agency – would therefore need to be frequently checked to ensure that it is functioning correctly. This is because a system malfunction could prevent people from accessing information that is crucial to making an independent decision.

PRIVACY

To safeguard human autonomy in the context of AI innovation in law enforcement, it is important that **agencies engage with AI systems in a way that protects the private sphere of individuals, including the users of the AI system, victims, suspects, and the general public.** This entails safeguarding their physical and mental integrity, personal relationships, personal space and home, and personal data in general, as this is essential for individuals to maintain their capacity to self-govern and exercise their rights.

WANT TO LEARN MORE?

See the [“Data protection in the digital age”](#) section in the annex.



Respecting privacy is a general condition of principled policing. By its very nature, law enforcement work requires the collection and analysis of information often related to the private lives of individuals. Therefore, the duty of confidentiality is a common element across the various professional rules for law enforcement officers. For similar reasons, domestic laws include safeguards governing when officers²² may justifiably interfere in someone’s private sphere – for example, requiring officers to obtain a warrant for a house search or seek legal authority to gather and retain any data that may be classified as being of a private nature. As

AI systems boost the potential for collecting and analysing information in general, and personal data in particular, law enforcement agencies should be particularly mindful of privacy in their AI-related activities.



Image by MohamadFaizal - stock.adobe.com

WANT TO LEARN MORE?

See the [“Privacy-by-design and Privacy Enhancing Technology”](#) section in the annex.



Respecting privacy is also fundamental to fulfilling the principle of lawfulness. Privacy is a human right that protects an individual’s private life, family life, home, and correspondence from arbitrary or unlawful interferences. Therefore, any interference with the right to privacy must be limited by the principles of legitimacy, necessity, and proportionality throughout the AI life cycle. Law enforcement agencies could benefit from concepts such as privacy-by-design and privacy-enhancing technology to facilitate the process of integrating the protection of human rights, including the right to privacy, into the development, procurement, and use of AI systems with intrusive potential. Importantly, law

enforcement agencies should be aware of the potential privacy concerns that may arise in relation to the training data used to build the AI system, putting in place sufficient safeguards to protect such data or requiring external developers to do so.

**PRACTICAL
EXAMPLE**
Safeguarding privacy in AI-enhanced surveillance

Surveillance operations are important for law enforcement agencies in detecting, investigating, and gathering evidence of crimes. However, requirements, processes and safeguards should be in place to ensure that surveillance is conducted within the law, including human rights law.²³ This becomes particularly relevant as AI systems are increasingly used to enhance the surveillance capabilities of law enforcement agencies, given the potential such systems have of amplifying risks to privacy and other human rights. In line with the principles of legitimacy, necessity and proportionality, AI-enhanced surveillance should be limited to situations where there is sufficient suspicion of criminal activity and agencies are unable to use less intrusive means to obtain information with similar importance for their mission.

AI systems can also include capabilities and use techniques that limit the potential impact on privacy of certain surveillance activities. For example, AI-enhanced surveillance technology such as object recognition for CCTV cameras can be developed in such a way that any personal information collected by the cameras is automatically anonymized (for instance, faces and licence plates are blurred).

When the use of real-time facial recognition technology is allowed in restricted places, it can be developed in ways that minimize its negative impact on human rights through processes such as black listing or safe listing. For instance, a facial recognition system used in an airport or train station could check people's faces against a list of wanted criminals, suspects, missing persons, victims of human trafficking, etc. and automatically delete or anonymize the faces of those who are not considered a "match".

TRANSPARENCY AND EXPLAINABILITY

Responsible AI innovation entails that the people that interact with AI systems have enough knowledge and understanding of the systems to safeguard their autonomy. This is especially relevant in a law enforcement setting given the nature and the impact of law enforcement work and can be achieved by following the principles of transparency and explainability. These are related but distinct principles: while transparency focuses on promoting good communication practices throughout the AI life cycle, explainability aims to allow individuals to understand how the system reaches its outcomes.

To ensure transparency, law enforcement agencies are advised to verify that the developers of their AI system (internal or external) disclose all the necessary information and documentation to its users. This applies regardless of whether users are officers, other agency personnel or third parties – for instance, travellers that use an AI-enabled border control system or people that communicate with a chatbot when calling an agency for assistance. This is a precondition of many other aspects of responsible AI innovation, such as accountability, human control and oversight, and the ability to monitor the robustness, safety, and accuracy of the AI system.

While the specific information and documentation that needs to be disclosed varies according to the parties involved, the context, and the applicable legislation, transparency in the AI context generally covers:

- the system's purpose and intended context of use;
- the most relevant decisions taken during the design and development of the system, such as the main characteristics of the training data set, the data sources, and potential data set limitations (whether it is accurate, up to date, or representative).
- the type of AI algorithms and their limitations;
- the data the system collects and shares.

▶ *Learn more about data requirements in the **Technical Reference Book**.*

Transparency also requires that the public, and specifically those directly affected by the use of an AI system in law enforcement, be informed that such a system is being used or has been used by law enforcement agencies.

The individuals affected by the use of an AI system should be aware that the system is or has been used and be able to request additional information about the system. This is essential to safeguard their capability to contest the outputs of the system, hold those in charge accountable, and exercise their human rights. This element of transparency is therefore closely connected with the principle of lawfulness. This is particularly true in the criminal justice context, where suspects, criminals, and victims need to be able to access information about the AI systems that have been used during a criminal investigation, for instance, as part of their right to a fair trial.

Adequately informing the public is an important step in fostering trust and confidence in society regarding the use of AI systems in law enforcement. Such trust is essential as it allows AI systems to be implemented in a smoother and more sustainable way, and ultimately allows law enforcement agencies to pursue their mission.²⁴

▶ *Find out more about the role of public trust and how it relates to organizational culture in the **Organizational Roadmap**.*

**PRACTICAL
EXAMPLE****Public Algorithm Registers**

Publicly accessible algorithm registers are an example of the way information regarding the use of AI systems can be shared with the public. Such registers have been introduced in a number of cities with the aim of informing citizens about important elements in the development of AI systems used by public entities, including the data sets used to train the systems and the measures that have been put in place to ensure the systems' robustness, safety, accuracy, and fairness.²⁵

If necessary, this practice could also be beneficial for law enforcement agencies, as it could increase public acceptance of their use of AI systems. Nonetheless, special care should be taken that the information provided to the public does not include sensitive policing information, the disclosure of which could compromise the work of law enforcement agencies.

**COMMON
QUESTION****Does transparency compromise law enforcement work?**

In law enforcement, it may be necessary to hide certain information from the wider public to avoid compromising investigations and to protect the AI systems used from exploitation and evasion by malicious actors: transparency does not entail communicating to the public information that could compromise law enforcement work.²⁶

The principle of transparency involves disclosing the information and documentation that is necessary and adequate in a specific context and in accordance with the applicable laws. In fact, transparency with the public typically consists of providing general information about the AI systems being used: not detailed technical information on the specific models, but simply what type of algorithm has been chosen. Similarly, it does not demand the disclosure of sensitive or confidential data, but rather general details about what type of data was used to train the AI system and what data it collects. This general information does not provide potential malicious actors with any more information than, for instance, the fact that there are CCTV cameras in certain streets.

Explainability allows individuals to understand how and why an AI system has reached a particular outcome. It is crucial that the AI systems deployed by law enforcement agencies are explainable so that the people that use these systems or are affected by them can make sense of and meaningfully react to their outputs. In other words, without explainability, law enforcement agencies will inevitably struggle to implement effective human control and oversight or ensure contestability. A lack of explainability also undermines individuals' ability to obtain redress in the case of harmful errors.

WANT TO LEARN MORE?

See the [“Difference between explainability and interpretability”](#) section in the annex.



Explainability can be a challenge with some of the most complex AI systems. Certain machine learning models are considered *black boxes* because they are too complex for humans to understand.

▶ *Learn more about understanding AI systems and the black box problem in the **Technical Reference Book** and the **Introduction to Responsible AI Innovation**.*

In response to this issue, the field of “Explainable AI” has emerged, which aims to ensure that even **when humans cannot understand ‘how’ an AI system has reached an output, they can at least understand ‘why’ it has produced that specific output.** This field distinguishes explainability in a narrow sense, as different from interpretability.

Using black box systems for high-stakes decisions such as those taken in criminal justice and law enforcement contexts is controversial. It has also been argued that in some specific scenarios, and when analysing tabular data, the performance of explainable models can be similar to that of black-box models such as neural networks.²⁷ However, when dealing with complex data (i.e., audio/speech and video/images) deep learning systems are the state-of-the-art: other ‘non-black-box’ solutions cannot achieve the level of accuracy necessary for use in real-world scenarios.²⁸

In the context of criminal investigations, the explainability of AI systems used to obtain or analyse evidence is particularly important. In fact, in some jurisdictions, criminal evidence obtained with the support of AI systems has been challenged in courts on the basis of a lack of understanding of the way the systems function.²⁹ While the requirements for evidence admissibility are different in each jurisdiction, a sufficient degree of explainability needs to be ensured for any AI system used to obtain and examine criminal evidence. This helps guaranteeing, alongside the necessary technical competencies, that law enforcement officers involved in investigations and forensic examinations have sufficient understanding of the AI systems used to be able to ascertain and demonstrate the validity and integrity of criminal evidence in the context of criminal proceedings.

4. FAIRNESS

Fairness is a crucial principle for both AI ethics and criminal justice, and requires an equitable distribution of burdens and benefits, and resources and opportunities between individuals as well as across society. In the context of responsible AI innovation, **fairness means that law enforcement agencies should ensure, throughout their engagement with AI systems, a just and non-discriminatory treatment of individuals and groups and a contribution to a more equitable society.** Stakeholder involvement is particularly relevant to achieving this

kind of fairness. ▶ *Learn more about identifying and engaging with stakeholders in the section “Responsible AI Innovation in Action Workbook”.*

This substantive aspect of fairness is supplemented by a procedural aspect, which requires that agencies safeguard **people’s ability to contest decisions supported by AI systems and to be compensated if such decisions are harmful to them.**

The principle of fairness is closely connected with the principle of lawfulness, and especially the instrumental principle of proportionality. In fact, the balancing exercise between the negative effects a certain measure causes on people’s rights and the legitimate goal pursued is also a reflection of the principle of fairness.

The following principles are instrumental to fairness:

- [Equality and non-discrimination](#)
- [Protection of vulnerable groups](#)
- [Diversity and Accessibility](#)
- [Contestability and Redress](#)

EQUALITY AND NON-DISCRIMINATION

Respecting equality and non-discrimination within AI innovation in law enforcement means ensuring equal treatment and opportunities for all stakeholders and refraining from unjustifiably discriminating against individuals or groups throughout the AI life cycle.

Equality and non-discrimination are especially important in the context of responsible AI innovation in law enforcement.

WANT TO LEARN MORE?

See the “[Direct and indirect discrimination, AI systems and law enforcement](#)” section in the annex.



Firstly, the fair treatment of individuals is a key aspect of principled policing and is linked to the human rights to equality and non-discrimination that law enforcement agencies are legally obliged to respect.³⁰ For instance, a law enforcement agency that aims to implement an AI chatbot to interact with the public needs to ensure that people can still reach the agency via alternative means, so that those with less knowledge of or access to technology are not excluded from exercising their rights.

Secondly, discrimination in a law enforcement context poses a significant threat to individuals and society. For example, discrimination may lead to the wrongful prosecution and unjustified punishment of certain individuals – and, consequently, actual criminals remaining undiscovered.

Furthermore, introducing AI systems may enhance the risk of discrimination as these systems are susceptible to amplifying human biases. In a context where discrimination has historically been an issue, there is a risk that historical law enforcement data will reflect the individual and institutional prejudices that have had a disproportionate impact on certain individuals and groups. These prejudices could then make their way into the AI system.

To cultivate responsible AI, **law enforcement agencies need to ensure that the AI systems they use are trained with data sets containing the appropriate quality and quantity of data and that any identifiable and discriminatory biases are removed. Any decisions taken in the design and development of the system that may have a negative, unfair, or disproportionate impact on certain individuals or groups also need to be considered.** ▶ *Learn more about how AI systems may embed human values in the Introduction to Responsible AI Innovation. Learn more about data requirements in the Technical Reference Book.*

PRACTICAL EXAMPLE

Inclusion as a way of mitigating the risk of discrimination throughout the AI life cycle

Setting up teams of designers, developers, and coders with a diverse representation of gender, age, ethnicity, disability, and other characteristics is a first step towards reducing the risk of discrimination emerging during the development of an AI system. The active and constructive inclusion of a variety of experiences in the design and development stages creates more insight and awareness of negative stereotypes and prejudice and improves the ability to mitigate the risk of reproducing them in AI systems.

Similarly, training the law enforcement personnel who use an AI system to identify and account for possible biases in its outputs is decisive in terms of protecting individuals from discrimination. In fact, any decisions that affect individuals and their rights should ultimately lie in the hands of law enforcement officers or other personnel. It is therefore crucial to empower humans to verify outputs and avoid being over-reliant on the system. It is important to include a variety of perspectives throughout the other stages of the AI life cycle, including in the teams that use and monitor the AI systems.

PROTECTING VULNERABLE GROUPS

To pursue fair AI innovation, **law enforcement agencies should pay particular attention and due consideration to those groups who are most vulnerable to and at risk of being disadvantaged by the use of specific AI systems. Safeguards should be put in place throughout the AI life cycle to mitigate the risks and enhance the benefits for these groups.**

Through their design, development, deployment and use, AI systems may have a disproportionately negative impact on certain groups due to their characteristics or other circumstances. For example, differences in the accuracy of AI systems often affect certain groups more than others, especially because certain groups are more susceptible to being misrepresented in the data sets that are used to train the systems.

These groups often include people who are at a higher risk of being subjected to unjustified discrimination. This varies from region to region, but usually consists of racial and ethnic minorities, children, women, members of the LGBTQIA+ community, people living with physical and mental disabilities or in poverty, and people with a lack of access to education, work, and community. This principle is therefore closely connected to the principle of non-discrimination. It goes further, however, as it aims to guide those that develop and use AI systems to do so in a way that ensures equal access and opportunities or benefits for vulnerable groups instead of harming them – whether such harm amounts to a violation of the right to equality, non-discrimination, or other human rights.

DIVERSITY AND ACCESSIBILITY

In the context of AI innovation in law enforcement, diversity and accessibility mean that AI systems should be built to be usable by a wide range of individuals and groups, regardless of age, gender, ability, or other characteristics.³¹ This means verifying that the systems that are developed, procured and deployed are designed in a user-centric way and account for the various characteristics and abilities that the end users may have.

Building inclusive systems is crucial whenever these systems have an impact on people accessing public goods, services, or advantages. This principle is thus particularly relevant when law enforcement agencies develop, procure or use AI systems that are intended for use by the general public, as diversity and accessibility in the systems' design will have a direct impact on societal fairness. In fact, like any other tool, AI systems can empower people, or they can disenfranchise them due to lack of accessibility.

The principle of diversity and accessibility supplements both equality and non-discrimination and protecting vulnerable groups by highlighting the need to pursue fairness throughout AI system users' experience, by designing universal and accessible AI systems that do not leave anybody behind.

**PRACTICAL
EXAMPLE****The importance of diversity and accessibility in speech processing in command-and-control centres**

Law enforcement command-and-control centres have evolved over the years with the adoption of a variety of AI systems that can provide support with handling calls for assistance from the public. Speech analysis is among the many AI applications that have been explored in this context, and it has the potential to rapidly collect and analyse the statements of those calling for assistance. Speech processing is being developed to be applied, for instance, in automated call handling, triage solutions, or with speech-to-text capabilities to aid the collation and timely recording of call data on incident logs.³²

If developed and deployed correctly, an AI speech processing system could help law enforcement agencies pursue their mission in a fairer way, by prioritizing resources more effectively and objectively and improving record-keeping, which is essential for contestability and redress.

To achieve this, agencies need to make sure that the system in question has been developed to account for the diversity within the population in the area where it is designed to be used. In practice, this means that the system needs to be developed in a way that makes it accessible by people with different accents, dialects, languages, and speech abilities (to give a few examples), with a training data set that adequately represents all these different categories. The AI system also needs to have inbuilt mechanisms for human control and oversight. An AI system may fail to recognize what people are saying when they speak in a language, dialect, or accent that the system has not been trained to recognize. This may also be the case due to external factors such as a noisy background or a nervous caller who does not articulate clearly. In those situations, the AI system must be programmed to rapidly transfer the call to a human operator to ensure that a person in need does not get locked out of the system. This is particularly true if the system is used for automated call handling or triage.

An AI system without measures to ensure diversity and accessibility, especially when used in a time-critical, sensitive setting such as this, would be highly susceptible to failure, misclassification, or dysfunctionality in the collection of data. This would put both fairness and lawfulness at stake by negatively affecting the human rights and well-being of the people who are denied equal access to the emergency services and may potentially expose them to greater harm at a time when they are in need of help. It may also result in an inefficient response and a less than optimal use of resources by the call handlers and dispatchers within the law enforcement agency.


CONTESTABILITY AND REDRESS

The principle of contestability means that law enforcement agencies should ensure that the necessary technological and organizational measures are in place to allow both users and those affected by decisions based on the output of an AI system to challenge these decisions. Contestability focuses on the ability to argue against AI-supported decisions. It is linked to human control and oversight, transparency and explainability as well as good governance and its

instrumental principles, in that all these principles are requisites to properly fulfilling the principle of contestability.

WANT TO LEARN MORE?

See the “[Right to an effective remedy](#)” section in the annex.



The principle of redress means that agencies should go one step further and ensure that, when AI-supported decisions have an unjust negative impact, those affected are able to formally seek redress through adequate and accessible processes. Upholding the principle of redress also relates to the human right to an effective remedy, and therefore to the principle of lawfulness.³³

It is inevitable that AI systems will fail in some situations, or that individuals will suffer due to the decisions taken based on an AI system’s output. To foster trust in the use of AI systems in law enforcement, it is essential that users and the people impacted by those decisions are reassured that they can challenge them and be compensated for any harm they may suffer as a result.


5. GOOD GOVERNANCE

Good governance consists of establishing policies, processes, and structures within an organization that enable it to uphold human rights, adequately manage collective resources, and respond to the needs of the people that the organization aims to serve.³⁴ **In the context of AI innovation in law enforcement, good governance means that agencies should aim to set up an overarching structure for audits and accountability and to foster a culture of responsible AI innovation.** ▶ *To read more about the role of organizational culture in responsible AI innovation, refer to the [Organizational Roadmap](#).*

Good governance, human rights and the rule of law are all mutually reinforcing: the principles of human rights and the rule of law serve as a guide for good governance, and good governance is essential to upholding human rights and the rule of law.

WANT TO LEARN MORE?

See the “[Rule of law](#)” section in the annex.



The principle of good governance runs through the responsible AI innovation framework as it is essential to achieving the core principles of lawfulness, minimization of harm, human autonomy and fairness, and the respective instrumental principles.

The following principles are instrumental to good governance:

- [Traceability and Auditability](#)
- [Accountability](#)

TRACEABILITY AND AUDITABILITY

Traceability and auditability allow law enforcement agencies to duly supervise the development and use of an AI system, and in particular to prevent, identify and resolve any negative consequences that might arise from its use.

Good governance in AI innovation in law enforcement calls for agencies to **set up requirements, procedures, and technical solutions to ensure that the decision-making processes of an AI system are traceable, including adequately documenting the decisions made during design, development and use that influence the outputs of the AI system.** During use, traceability involves tracking and documenting AI outputs, including the input data used, the model and parameters selected, the model's output, the user's name, date, and any other relevant information. Traceability is important because it enables accountability and transparency, allowing stakeholders to understand how decisions were made and to identify any errors in the decision-making process.

In addition, law enforcement agencies should ensure that the AI systems they use are **auditable, in that their essential elements can be assessed by internal or external auditors.**

Ensuring that the inner workings of the AI system are traceable and can be assessed from the outside makes it easier for the principles of transparency, contestability, and redress to be fulfilled, and is central to evaluating the AI system's outputs and identifying and fixing any potential issues. Traceability and auditability should therefore be pursued and maintained throughout the AI life cycle, from conceptualization to monitoring.

ACCOUNTABILITY

When AI systems are used for decision-making processes in a law enforcement setting, it is crucial that **mechanisms are put in place to enable stakeholders to clearly determine who is responsible for the decisions made with the support of the AI system, and the consequences of those decisions.**

The central role that accountability plays in this context relates to the prominence of law enforcement in the functioning of society, justice, and governments, and consequently the high stakes for everyone involved. Because of the authority accorded to law enforcement agencies and officers, which is essential for the pursuit of their mission, there is an inherent power imbalance between those in charge of law enforcement and the rest of society. The complexity of AI systems, combined with the general population's lack of understanding of AI, could exacerbate this power imbalance when these systems are introduced. Responsible AI innovation compensates for this imbalance by requiring processes to be put in place to clearly determine which individuals are accountable for AI-related decisions.

Putting the principles into practice

The principles for responsible AI innovation are relevant throughout the AI life cycle. They aim to provide law enforcement agencies with an ethical and human rights-compliant way to navigate the many complex and crucial decisions that need to be taken, from the conceptualization to use and monitoring – and, in some cases, the decommissioning – of an AI system.

To put these principles into practice, it is helpful for agencies to follow a process of understanding and applying the principles, identifying and engaging with the relevant stakeholders, checking the results, and restarting if necessary. There is no set order, as the most appropriate way of performing each of these steps will vary depending on the circumstances. As illustrated in the figure below, this process should be followed throughout the AI life cycle and repeated cyclically.

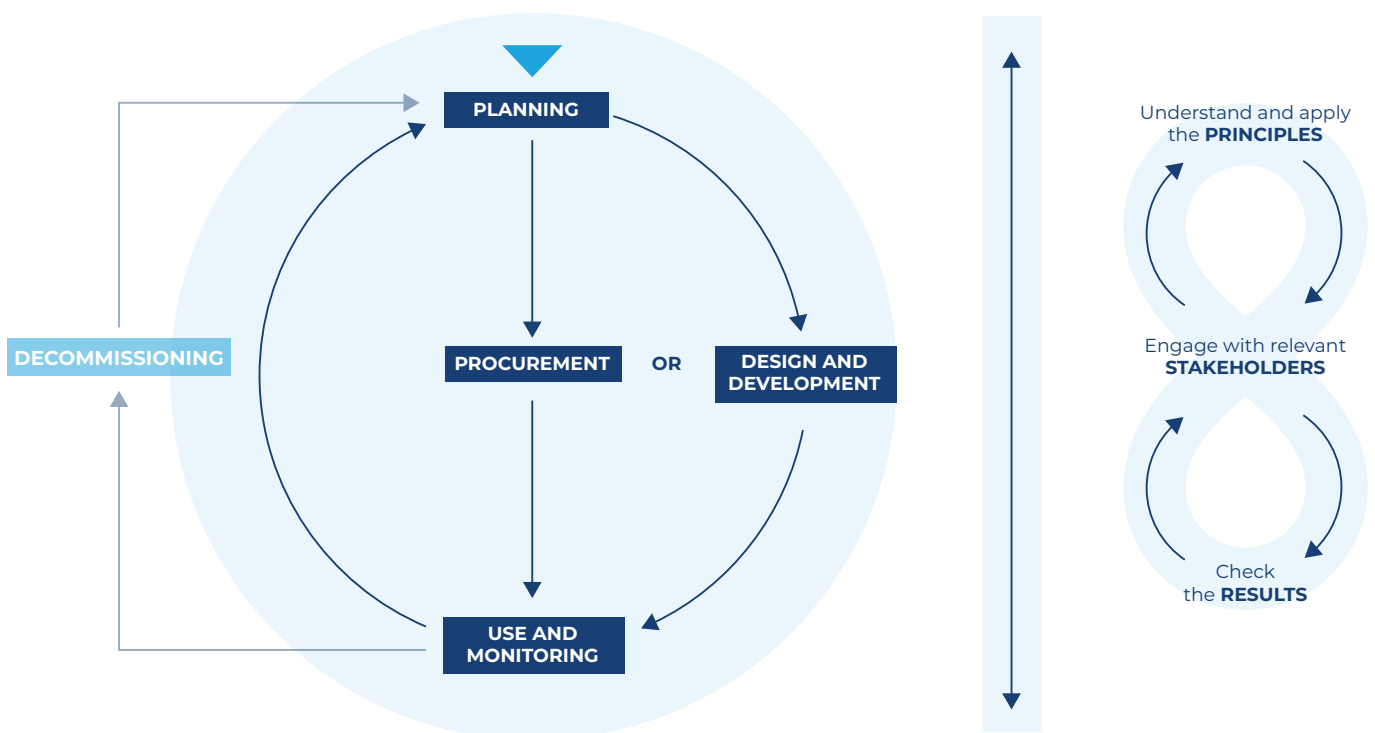


Figure 2 - Putting the principles into practice

In a nutshell, each of these steps entails the following:

UNDERSTANDING AND APPLYING THE PRINCIPLES

Law enforcement officers in their various relevant capacities are recommended to have a good understanding of the principles from the beginning of their engagement with an AI system. This document can be used as a basis to be consulted at any time to refresh or broaden one's knowledge of each of the principles.

The principles are meant to be followed throughout the AI life cycle to support all decision-makers in a law enforcement agency in evaluating the impact of an AI system on individuals, society, and the environment, and establishing the measures that can be taken to avoid or mitigate any negative consequences.



Image by Wasan - stock.adobe.com

In practice, this involves asking different questions at each stage, thus allowing agencies to thoroughly explore and address the positive and negative consequences of implementing any given AI system. ▶ *Learn more about how the principles translate into each stage of the AI life cycle in the **Responsible AI Innovation in Action Workbook**.*

As noted at the beginning of this document, the instrumental principles may sometimes be conflicting, or law enforcement agencies may not be able to fulfil each of them to their full potential. This means that trade-offs may be necessary, and it is important that law enforcement officers are well equipped to make the appropriate decisions and to document/record the decisions made. However, responsible AI innovation requires the core principles to be upheld at all times. For instance, when developing a certain AI system, a decision may need to be made between maximizing either accuracy or explainability, or either privacy or transparency; whatever the decision, the law, including human rights law, must be respected.

IDENTIFYING AND ENGAGING WITH THE RELEVANT STAKEHOLDERS

There is an increased expectation from workers, criminal justice practitioners, regulators, and society in general that they will be involved in high stakes decisions related to AI innovation in law enforcement. Successfully implementing new AI systems in an agency therefore requires **identifying and engaging with the relevant stakeholders**. This also applies to AI systems already in use in the various agencies. In other words, law enforcement agencies are advised to carefully determine those who might have a stake in the implementation of the AI system and involve them in the process as appropriate. In the context of law enforcement, these stakeholders may include:

- the individuals who are subject to and may benefit from or be harmed by the use of an AI system, such as suspects, victims, civil society groups, and the general public
- the individuals whose data is used to test and develop AI systems
- innovation units and development teams both within law enforcement and in the private sector who develop AI systems and tools
- law enforcement officers and other personnel who interact with AI systems
- law enforcement management, who will be accountable for deploying an AI system too early or for missing an opportunity to use an AI system
- practitioners within criminal justice systems who need to make sense of the information and decisions that they receive from law enforcement

▶ *Learn more about how to identify the stakeholders in the **Responsible AI Innovation in Action Workbook**.*

Once these stakeholders have been identified, their perspectives, concerns, and goals should be understood. This can be achieved in many ways, depending on the case. For instance, the individuals who are subject to the use of an AI system could be involved through consultation sessions or a review of high-quality research on the topic. Law enforcement officers who will interact with AI systems could be involved through feedback sessions or training courses. ▶ *Learn more about specific ways of engaging with the general public in the **Organizational Roadmap**.*

CHECKING THE RESULTS

As law enforcement agencies advance through the AI life cycle, they should keep in mind the principles and the relevant stakeholders. Agencies are advised to keep track of the consequences of their decisions and the results of their activities, and correct their course if needed.

REPEAT (IF NEEDED)

After checking the results, law enforcement agencies may find that they need to re-evaluate or re-interpret the principles for responsible AI innovation, identifying different stakeholders or engaging with them in different ways. AI innovation in law enforcement is always evolving, but a proper understanding of the principles and adequate interaction with stakeholders will allow agencies to move ahead in a responsible manner.

Annex:

Want to learn more?

1. THE FOUNDATIONS OF THE CORE AND INSTRUMENTAL PRINCIPLES

The principles for responsible AI innovation are anchored in fundamental concepts from ethics and human rights law and are aligned with policies, regulations, and principles relevant to AI and policing that have been established at a national, regional, and international level.

Firstly, the concepts on AI are based on the following sources:

- Recommendations on the Ethics of Artificial Intelligence adopted by the General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), meeting in Paris from 9 to 24 November 2021, at its 41st session.
- AP4AI Framework Blueprint issued in the context of the project Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain, coordinated by Europol and CENTRIC.³⁵
- Ethics Guidelines for Trustworthy AI by the High-Level Expert Group set by the European Commission.³⁶

Secondly, concepts referring to human rights are derived from the Universal Declaration of Human Rights and the main human rights treaties. Lastly, principles on policing are built upon the Code of Conduct for Law Enforcement Officials adopted by the United Nations General Assembly³⁷ and the Peelian policing principles.

You can learn more about the ethics and human rights foundations of the principles in the *Introduction to Responsible AI Innovation*. In a nutshell, the core principle of lawfulness reflects a basic principle of good conduct on the part of law enforcement officials and encompasses agencies' and officials' general obligation to respect human rights law. The core principles of minimization of harm, human autonomy and fairness correspond to basic principles of ethics which are grounded in various philosophical theories that argue for the inherent value of each of these principles. The core principle of good governance also draws on human rights law and ethics and relates to the overarching structures that are needed to achieve responsible AI innovation.

2. HUMAN-IN-THE-LOOP, HUMAN-ON-THE-LOOP, HUMAN-IN-COMMAND

Human oversight helps ensure that an AI system does not undermine human autonomy or cause any other adverse effects. One way this can be achieved is through oversight mechanisms that place a “human-in-the-loop”, thus allowing for human intervention at every decision cycle in the AI system’s development and use.

As this may be too burdensome in certain contexts and use cases involving AI systems, an alternative mechanism is the so-called “human-on-the-loop”, whereby human intervention is guaranteed during the design of the AI system and while monitoring it during use.

A third possible approach is integrating a “human-in-command” mechanism. This entails the capability of humans to oversee not only the overall activity in the AI system, but also its impact on groups of people, societal or economic structures, or legal obligations, and to decide when and how to use the AI system.³⁸

3. DATA PROTECTION IN THE DIGITAL AGE

In recent years, a significant number of countries across the world have introduced data protection laws to regulate the processing of personal data by state authorities, businesses, and other actors.³⁹ The protection of personal data ensures the integrity and confidentiality of data, and provides the person, the so-called data subject, with a right to control their data. This is typically reflected in a right to be informed about the collection, processing, storage, and sharing of data by other actors, for example public authorities or companies.

Domestic laws and regulations set up criteria for data processing, including purpose specificity, data minimization, and storage limitation, and for processing security to ensure the confidentiality, integrity, and availability of data. They also set out the responsibilities of suppliers who process data on behalf of another actor who has collected the data, typically referred to as the data controller. The data controller is responsible for establishing the purpose and the legal basis for the processing of data.

Privacy laws differ between countries, but the human right to privacy and the ethical imperative to respect privacy concerns remain, regardless of the specific domestic laws. This is receiving increased attention in relation to new technology such as AI systems.

4. PRIVACY-BY-DESIGN AND PRIVACY-ENHANCING TECHNOLOGY

As societies have become more aware of the intrusive potential of certain technology, concepts such as *privacy-by-design* and privacy enhancing technology have increasingly gained prominence. *Privacy-by-design* refers to the process of embedding data protection and privacy considerations from the start of the creation process for a piece of technology, to ensure that privacy is protected throughout the technology's life cycle.⁴⁰ *Privacy-enhancing technology* is technology that incorporates techniques allowing information to be processed while protecting confidentiality and upholding privacy, such as encryption, data anonymization and federated learning.

5. WHAT IS THE DIFFERENCE BETWEEN EXPLAINABILITY AND INTERPRETABILITY?

In the context of AI, interpretability and explainability (in a narrow sense) are related but distinct concepts. Typically, these terms are used with the following definitions:

- **Explainability** (in a narrow sense) refers to the ability of developers and users of an AI system to understand its functioning, meaning how the system makes decisions or generates outputs. It focuses on the inner workings of the AI system, its internal logic, or underlying processes.
- **Interpretability**, on the other hand, refers to the ability to provide reasoning for a specific outcome the system has produced – in other words, to understand why a certain result has been generated.⁴¹

Several techniques are being developed in the field of “Explainable AI” that aim to ensure the interpretability of non-explainable models. For example, in an object recognition task, “Explainable AI” requires identifying which pixels or parts of the image have led to a specific output, without necessarily understanding the full path the AI system has taken from input to output.

6. DIRECT AND INDIRECT DISCRIMINATION, AI SYSTEMS AND LAW ENFORCEMENT

Equality and non-discrimination play a prominent role both as fundamental principles of international human rights law and human rights in themselves. The rights to equality and non-discrimination are recognized in the Universal Declaration of Human Rights as well as in many international and regional treaties, some general and some specifically focused on eliminating certain forms of discrimination. They impose on states and state bodies the obligation to ensure that individuals are treated equally and are equally able to exercise and enjoy all their rights, and that everyone is protected against direct or indirect discrimination based on “protected characteristics” such as gender, race, ethnic origin, age, religion, ability, and sexual orientation.

Discrimination has long been an issue in the law enforcement context across the world. In 2021, the United Nations Committee on the Elimination of Racial Discrimination released its findings regarding several countries in Europe, Asia, and South America, and expressed concerns including the excessive use of force against certain ethnic groups, high numbers of instances of racial hate speech, ethnicity-based facial recognition which may lead to racial profiling, and a high proportion of members of ethnic minorities awaiting death sentences.⁴² Similarly, a study by the European Union Agency for Fundamental Rights showed that some ethnic communities are more likely to be stopped by the police in Europe.⁴³ In some countries, the issue of over-policing and the use of excessive force in certain communities is widely researched and a recent analysis found that minority communities are twice as likely to be fatally shot by police than majority communities.⁴⁴

Direct and indirect discrimination in law enforcement can take various forms, each manifesting itself differently and thus requiring different mitigation techniques. For instance, while racial bias often causes some communities to be more heavily policed and punished, gender bias can take the form of dismissal or neglect of women reporting domestic or partner violence and mistreatment of members of the LGBTQAI+ community.⁴⁵

In the context of AI, it is important to note that even when the data that is used to train the systems or that is processed by them does not refer to “protected characteristics” under the right to non-discrimination, other categories may be used as proxies for protected categories. For example, a category may not directly address race or gender, but it may incidentally reveal such information.

7. RIGHT TO AN EFFECTIVE REMEDY

The right to an effective remedy reflects an obligation on the part of public authorities to set up complaint mechanisms allowing individuals to submit complaints regarding restrictions or violations of human rights. This right is an integral element of human rights law. It is an essential safeguard in providing effective recourse to anyone who alleges that their rights have been interfered with. Without such recourse, human rights cannot be fully exercised and enjoyed.

The United Nations Human Rights Committee has stressed the need for establishing remedies “to give effect to the general obligation to investigate allegations of violations promptly and effectively through independent and impartial bodies.” The Committee also highlights the necessity for remedies to be not only effective and enforceable but also accessible and appropriately adapted to the needs of groups of persons with vulnerabilities.⁴⁶

International and regional human rights conventions include various measures aimed at ensuring effective remedies, in provisions regarding the right to a fair trial or specific rights to judicial protection and access to the courts. Such complaints mechanisms should be determined by judicial, administrative, or legislative authorities at a national level.

8. THE RULE OF LAW

In simple terms, the rule of law means that every person and every entity, both private and public, is accountable to the law – including the State and state officials. It means that everyone is equally obliged to follow laws that are enacted in a public and independent manner and that are enforced in a fair manner, in accordance with international standards.⁴⁷

Ensuring the rule of law presupposes the separation of powers, participation in decision-making, legal certainty, the avoidance of arbitrariness, and procedural and legal transparency. This includes a judicial system which is accessible and impartial and delivers timely decisions made by competent, ethical, and independent representatives.⁴⁸

The principles of the rule of law pave the way for people’s access to public services, curbing corruption, restraining the abuse of power, and safeguarding the social contract between people and the state. As such, the rule of law is an essential foundation of fair, stable and cooperative relations between countries and within countries, which foster social progress and development.

ENDNOTES

- 1 United Nations Office on Drugs and Crime. (2006). Compendium of United Nations standards and norms in crime prevention and criminal justice. P. xi. Accessible at: https://www.unodc.org/pdf/criminal_justice/Compendium_UN_Standards_and_Norms_CP_and_CJ_English.pdf
- 2 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials. Accessible at: <https://digitallibrary.un.org/record/10639?ln=en#record-files-collapse-header>
- 3 Cansu Canca (2020). Operationalizing AI Ethics Principles. Communications of the ACM, December 2020, Vol. 63 No. 12, Pages 18-21. Accessible at <https://cacm.acm.org/magazines/2020/12/248788-operationalizing-ai-ethics-principles/abstract>
- 4 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials, art 8. Accessible at: <https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials>
- 5 Europol and CENTRIC. (2022). Accountability Principles for AI (AP4AI) in the Internal Security Domain: AP4AI Framework Blueprint, p. 27-28. Accessible at: https://www.ap4ai.eu/sites/default/files/2022-03/AP4AI_Framework_Blueprint_22Feb2022.pdf
- 6 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials, art. 2. Accessible at: <https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials>
- 7 United Nations Office of the High Commissioner for Human Rights (OHCHR). (2006). Frequently Asked Questions on A Human Rights-Based Approach to Development Cooperation, p. 5. Accessible at: <https://www.ohchr.org/sites/default/files/Documents/Publications/FAQen.pdf>
- 8 United Nations Human Rights Council. (2015). Use of information and communications technologies to secure the right to life A/HRC/26/36 §56. Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns. Accessible at: <https://daccess-ods.un.org/tmp/3125799.59630966.html>
- 9 United Nations Human Rights Committee. (1999). CCPR General Comment No. 27: Article 12 (Freedom of Movement). Accessible at: <https://digitallibrary.un.org/record/366604?ln=en>
- United Nations Human Rights Committee. (2020). CCPR General Comment No. 37: Article 21 (Right of peaceful assembly). Accessible at: <https://daccess-ods.un.org/tmp/3062642.51470566.html>
- 10 Amnesty International. (2015). Use of Force: Guidelines for implementation of the UN basic principles of the use of force and firearms by law enforcement officials. Amnesty International, Dutch Section. Police and Human Rights Programme. Accessible at: https://www.amnesty.org.uk/files/use_of_force.pdf
- 11 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials., art 3. Accessible at: <https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials>
Basic Principles on the Use of Force and Firearms by Law Enforcement Officials, Adopted by the Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Havana, Cuba, 27 August to 7 September 1990. Accessible at: <https://www.ohchr.org/sites/default/files/firearms.pdf>
- Amnesty International. (Aug. 2015). Use of Force: Guidelines for implementation of the UN basic principles of the use of force and firearms by law enforcement officials. Amnesty International, Dutch Section. Police and Human Rights Programme. Accessible at: https://www.amnesty.org.uk/files/use_of_force.pdf

- 12 European Court of Human Rights. (Dec. 1976). Case of Handyside v. United Kingdom, No.5493/72, para. 48-49. Council of Europe. Accessible at: [https://hudoc.echr.coe.int/eng#{"dmdocnumber":\["695376"\],"itemid":\["001-57499"\]}](https://hudoc.echr.coe.int/eng#{)
- 13 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials., art. 2. Accessible at: <https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials>
- 14 High-Level Expert Group on AI set by the European Commission. (Apr. 2019). Ethics guidelines for trustworthy AI, pp.16-17. Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 15 National Institute of Standards and Technology (NIST), U.S. Department of Commerce. (n.d.). Safety – Glossary (Computer Security Resource Center). Accessible at: <https://csrc.nist.gov/glossary/term/safety>
- 16 High-Level Expert Group on AI set by the European Commission. (Apr. 2019). Ethics guidelines for trustworthy AI, p.17. Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 17 World Economic Forum, UNICRI, INTERPOL and Netherlands Police. (2022). A Policy Framework for Responsible Limits on Facial Recognition Use Case: Law Enforcement Investigations. Accessible at: <https://unicri.it/sites/default/files/2022-11/A%20Policy%20Framework%20for%20Responsible%20Limits%20on%20Facial%20Recognition.pdf>
- 18 Jacqueline G. Cavazos, P. Jonathon Phillips, Carlos D. Castillo and Alice J. O’Toole. (Jan. 2021). "Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?". IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 101-111, Accessible at: <https://doi.org/10.1109/TBIOM.2020.3027269>;
- 19 Patrick Grother, Mei Ngan, Kayee Hanaoka. (Nov, 2022). Face Recognition Vendor Test (FRVT), Part 2: Identification. Information Access Division, Information Technology Laboratory. NIST Interagency Report 8271. Accessible at: https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf
- 20 High-Level Expert Group on AI set by the European Commission. (Apr. 2019). Ethics guidelines for trustworthy AI, p.16. Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 21 Europol and CENTRIC. (2022). Accountability Principles for AI (AP4AI) in the Internal Security Domain: AP4AI Framework Blueprint, p. 84. Accessible at: https://www.ap4ai.eu/sites/default/files/2022-03/AP4AI_Framework_Blueprint_22Feb2022.pdf
- 22 United Nations General Assembly (1979-1980). Resolution adopted by the General Assembly on 17 December 1979. 34/169. Code of Conduct for Law Enforcement Officials. Accessible at: <https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials>
- 23 International Association of Chiefs of Police (IACP). (Apr. 2009). Surveillance. Law Enforcement Policy Center. Accessible at: <https://www.theiacp.org/sites/default/files/2020-06/Surveillance%20FULL%20-%2006222020.pdf>
- 24 Europol and CENTRIC. (2022). Accountability Principles for AI (AP4AI) in the Internal Security Domain: AP4AI Framework Blueprint, p. 30. Accessible at: https://www.ap4ai.eu/sites/default/files/2022-03/AP4AI_Framework_Blueprint_22Feb2022.pdf
- 25 Algorithm Register. (n.d.). Algorithmic Transparency Standard. Accessible at: <https://www.algorithmregister.org/> <https://www.algorithmregister.org>
- 26 Europol and CENTRIC. (2022). Accountability Principles for AI (AP4AI) in the Internal Security Domain: AP4AI Framework Blueprint, p. 31. Accessible at: https://www.ap4ai.eu/sites/default/files/2022-03/AP4AI_Framework_Blueprint_22Feb2022.pdf

- 27 Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, & Cynthia Rudin. (2018). Learning Certifiably Optimal Rule Lists for Categorical Data (arXiv:1704.01701). arXiv. <https://doi.org/10.48550/arXiv.1704.01701>
- Jiaming Zeng, J., Berk Ustun, & Cynthia Rudin. (2017). Interpretable Classification Models for Recidivism Prediction. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(3), 689–722. <https://doi.org/10.1111/rssa.12227>
- Jacqueline G. Cavazos, P. Jonathon Phillips, Carlos D. Castillo and Alice J. O’Toole. (Jan. 2021). Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 1, pp. 101-111, Accessible at: <https://doi.org/10.1109/TBIOM.2020.3027269>;
- Haiyu Wu, Vitor Albiero, K. S. Krishnapriya, Michael C. King, Kevin W. Bowyer. (2022). Face Recognition Accuracy Across Demographics: Shining a Light Into the Problem. arXiv:2206.01881. Accessible at: <https://doi.org/10.48550/arXiv.2206.01881>
- Tollenaar, N., & van der Heijden, P. G. M. (2013). Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 565–584. <https://doi.org/10.1111/j.1467-985X.2012.01056.x>
- 28 Pantelis Linardatos, Vasilis Papastefanopoulos, & Sotiris Kotsiantis. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1), Article 1. <https://doi.org/10.3390/e23010018>
- 29 George Washington University. (n.d.). AI Litigation Database. Accessible at: <https://blogs.gwu.edu/law-eti/ai-litigation-database/>
- 30 United Nations Human Rights Committee, General comment No. 18: Non-discrimination, Thirty-seventh session (1989).
- 31 European Union Independent High-Level Expert Group on Artificial Intelligence. (2019, April 8). Ethics guidelines for trustworthy AI | Shaping Europe’s digital future, p.18. Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 32 Andy Doran. (Aug. 2020). Answering the call: Lancashire’s voice to text analytics project. *Policing Insight*. Accessible at: <https://policinginsight.com/features/innovation/answering-the-call-lancshires-voice-to-text-analytics-project/>
- 33 United Nations. (n.d.). Universal Declaration of Human Rights, Art. 8. United Nations. Accessible at: <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- 34 United Nations Office of the High Commissioner for Human Rights (OHCHR). (n.d.). About good governance. OHCHR and good governance. Accessible at: <https://www.ohchr.org/en/good-governance/about-good-governance>
- 35 Accessible at: https://www.ap4ai.eu/sites/default/files/2022-03/AP4AI_Framework_Blueprint_22Feb2022.pdf
- 36 Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 37 Accessible at: <https://digitallibrary.un.org/record/10639?ln=en#record-files-collapse-header>
- 38 High-Level Expert Group on AI set by the European Commission. (Apr. 2019). Ethics guidelines for trustworthy AI. Accessible at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 39 For example: Brazil. Data Protection Law (LGPD). Law No. 13,709, of August 14, 2018, (Wording given by Law No. 13.853 of 2019); United States, California. California Consumer Privacy Act of 2018 (CCPA); People’s

Republic of China. Personal Information Protection Law (PIPL). European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). South Africa. Protection of Personal Information Act (POPIA). Act No. 4 of 2013.

- 40 United Nations Development Group. (n.d.). Data Privacy, Ethics and Protection. Guidance Note on Big Data for Achievement of the 2030 Agenda. Accessible at: https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf
- 41 Pantelis Linardatos, Vasilis Papastefanopoulos, & Sotiris Kotsiantis. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1), Article 1. <https://doi.org/10.3390/e23010018>
- 42 United Nations Committee on the Elimination of Racial Discrimination. 105th Session (15 Nov 2021 - 03 Dec 2021). Findings on Chile, Denmark, Singapore, Switzerland and Thailand. Accessible at: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/SessionDetails1.aspx?SessionID=2484&Lang=en
- 43 European Union Agency for Fundamental Rights (FRA). (May 2021). Police stops in Europe: Everyone has a right to equal treatment. European Union Agency for Fundamental Rights. Accessible at: <http://fra.europa.eu/en/news/2021/police-stops-europe-everyone-has-right-equal-treatment>
- 44 Lynne Peeples. (2019). What the data say about police shootings. *Nature*, 573(7772), 24–26. <https://doi.org/10.1038/d41586-019-02601-9>
- 45 Michele Decker, Charvonne N. Holliday, Zaynab Hameeduddin, Roma Shah, Janice Miller, Joyce Dantzer, & Leigh Goodmark. (2019). You Do Not Think of Me as a Human Being: Race and Gender Inequities Intersect to Discourage Police Reporting of Violence against Women. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 96(5), 772–783. Accessible at: <https://doi.org/10.1007/s11524-019-00359-z>
- 46 United Nations Human Rights Committee. (26 May 2004). General comment no. 31 [80], The nature of the general legal obligation imposed on States Parties to the Covenant. CCPR/C/21/Rev.1/Add.13. Accessible at: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2FC%2F21%2FRev.1%2FAdd.13&Lang=en
- 47 United Nations (n.d.). What is the Rule of Law. United Nations and the Rule of Law. Accessible at: <https://www.un.org/ruleoflaw/what-is-the-rule-of-law/>
- 48 World Justice Project. (n.d.). What is the Rule of Law? Accessible at: <https://worldjusticeproject.org/about-us/overview/what-rule-law>

How to cite this publication: UNICRI and INTERPOL. (Revised February 2024).

Toolkit for Responsible AI Innovation in Law Enforcement: **Principles for Responsible AI Innovation.**

© United Nations Interregional Crime and Justice Research Institute (UNICRI), 2024

© International Criminal Police Organization (INTERPOL), 2024



www.interpol.int
www.unicri.it



INTERPOL_HQ



@INTERPOL_HQ
@UNICRI



INTERPOL HQ
UNICRI



INTERPOL
UNICRI



@INTERPOL
@UNICRIHQ

www.ai-lawenforcement.org