A European Positive Sum Approach towards AI tools in support of Law Enforcement and safeguarding privacy and fundamental rights

D2.4: Ethical frameworks for the use of LEAs

Grant Agreement ID	101022001	**Acronym**	PopAI
Project Title	A European Positive Sum Approach towards AI tools in support of Law Enforcement and safeguarding privacy and fundamental rights		
Start Date	01/10/2021	**Duration**	24 Months
Project URL	https://www.pop-ai.eu/		
Document date	21/02/2023		
Nature	R = Document, report	**Dissemination Level**	PU = Public
Authors	Francesca Trevisan (Eticas), Carlos Zednik (TU/e)		
Contributors	Pinelopi Troullinou (TRI), Xenia Ziouvelou (NCSRD), Sofia Segkouli (ITI), Dimitra Papadaki (KEMEA), Isabela Miranda (Eticas)		
Reviewers	Dimitris Kyriazanos, Andreas Ikonomopoulos (NCSRD)		

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101022001.

Executive Summary

Law enforcement agencies and judicial authorities around the world are turning to AI, making it an important tool for crime prevention, investigation, and training among other areas. However, AI also raises important concerns about transparency, accountability, and respect for human rights. The need for democratic oversight of AI is growing as evidence mounts of potential misuse and infringement of rights. The lack of trust around AI in security is due to AI's socio-technical limitations, its opacity, and the power imbalance between those using the technology and those who are subjected to it.

To foster trust in AI and for a more efficient development and use of AI in law enforcement, it is key to understand what ethics is, why it is important and how it applies to AI in the field of law enforcement. To achieve this, this report explains what ethics and AI ethics are, their key mechanisms, and the main ethical concerns for AI applications in law enforcement. Furthermore, it organises the current ethical frameworks in the LEA and AI space in a systematic and extensible taxonomy. Finally, it presents some novel ethics guidelines for AI in Law enforcement. These guidelines include practical recommendations on how to: 1) ensure and promote public safety in the digital domain, 2) restrain AI intervention, 3) avoid AI bias and promote impartiality, 4) act with integrity and respect for the rule of law when using AI, 5) protect and promote individual rights when using AI, and 6) apply community policing to AI use.

Table of Contents

List of Figures

List of Terms & Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AI HLEG	High Level Expert Group on Artificial Intelligence
D	Deliverable
DPA	Data Protection Authority
DPIA	Data Protection Impact Assessment
HRIA	Human Right Impact Assessment
LEAs	Law Enforcement Agencies
UN	United Nations
VR	Virtual Reality
WP	Work Package

1 Introduction

Artificial intelligence (AI) has entered the law enforcement sector with the promise of bringing more efficiency and effectiveness. However, AI came with a number of risks to human rights and democracies caused by several factors such as its rapid progress, opaqueness, bias, and the potential for misuse by individuals and organizations (see D2.3 and D3.1). The use of AI in policing can also erode trust between communities and law enforcement, leading to further social unrest (see D2.3).

In order to maximise the benefits of AI for society, it's important to implement proper regulations, and safeguards to prevent and mitigate AI risks. While current laws and regulatory proposals applying to AI have been analyzed and classified in Deliverable 2.2, this report focuses on another crucial aspect of risk prevention and management: ethics.

A culture of AI ethics maximises AI benefits by ensuring that the technology is developed, deployed and used in a responsible manner. To create a sustainable and inclusive European AI hub for LEAs, it is important to understand what ethics is, what it wants to achieve, why it is important and how it applies to LEAs use of AI. To this end, this report systematizes the knowledge around ethics, AI and law enforcement and develops new practical guidelines for LEAs' responsible use of AI.

1.1 Scope and objectives of the deliverable

This deliverable pertains to the work conducted within Work Package (WP) 2 “Security AI in the next 20 years: trends, practices and risks”. WP2 builds on the existing state of the art in relation to the use of AI by LEAs in Europe and elsewhere to identify:

1) the actual AI use and technical characteristics of AI tools in the security domain (Task 2.1);

2) the legal frameworks and recent court rulings (Task 2.2);

3) how controversies have shaped technology adoption in the security domain (Task 2.3);

4) the ethical principles and challenges that can inform a practical ethics toolbox (Task 2.4);

5) the organisational issues around AI implementation in LEA contexts (Task 2.5).

Task 2.4 "The Ethical Principles and Challenges that can Inform a Practical Ethics Toolbox" explores current ethical frameworks in the LEA and AI space, drawing on published materials and internal reports used by security actors and technology providers. Its findings are divided into two deliverables: D2.4: "Ethical frameworks for the use of AI by LEAs" and D2.5 "Practical ethics toolbox for the use of AI by LEAs". This report covers D2.4 and systematises ethics knowledge on AI and LEAs to highlight the main common points between frameworks and identify key gaps. This report wants to achieve four primary goals:

- Define the concept of ethics and its relevance to AI applications in law enforcement
- Highlight the ethical risks associated with various AI applications in policing

- Help LEAs navigate the complex field of AI ethics for LEAs by creating a taxonomy
- Propose a set of guidelines for the responsible use of AI by LEAs.

1.2 Structure of the deliverable

To achieve the objectives outlined above, this deliverable is organised into five main sections.

Section 1 introduces the main topic discussed in the deliverable, outlines its scope and explains how this work relates to other popAI tasks and deliverables.

Section 2 defines ethics, AI ethics and the main ethics mechanisms. It presents the objectives of ethics, it stresses its importance, and it analyses the key ethical concerns regarding the use of AI for crime prevention and investigation, migration management, administration of justice, cyber operations, and LEAs' training.

Section 3 outlines the methodology employed to develop the popAI taxonomy and guidelines. It discusses in details the key ethics frameworks and principles relevant to law enforcement agencies and AI and shows the gaps.

Section 4 lays out the popAI guidelines, which have been created through a combination of existing LEA guidelines and AI principles. The guidelines are accompanied by illustrative examples drawn from the popAI functionalities taxonomy (D2.1) and controversies report (D2.3), and by key recommendations that ease their implementation.

Finally, section 5 provides a discussion and some key recommendations.

1.3 Relation to other tasks and deliverables

This report contributes to the popAI project by organising the material that has been produced on ethics in the areas of AI and law enforcement, by emphasizing crucial gaps and future directions and by providing a list of key ethical guidelines for AI in law enforcement. Organising in a systematic manner the knowledge around AI and LEAs ethics is key to assist LEAs, policymakers, researchers and other stakeholders to navigate this field, take decisions on policies, organizational strategies, research and training.

This report adds an ethical perspective to the popAI functionalities (D2.1) and legal (D2.2) taxonomies and provides the background for the popAI ethics toolbox (D2.5). Furthermore, this work supports WP3 empirical studies by providing key ethics references. Finally, the recommendations made by WP4 to civil society, LEAs, and technology developers will be based on the ethical frameworks and gaps documented in this report.

1.1 Work methodology

This report was generated through a comprehensive examination of both academic and non-academic sources, with contributions from all PopAI research collaborators. The sources of information can belong to broader five categories:

• Scientific Publications: Collection of publications in scientific journals, conference/workshops

proceedings and scientific book chapters.

• Gray literature: magazines and newspaper articles.

• Official report: work published by public and private institutions in a report format.

• Legal Frameworks: Information related to regulations that regard technology and innovation.

• Advocacy work: work produced by organizations for advocacy purposes.

Additionally, we have sought the input of crucial project entities (ALIGNER and STARLIGHT), who offered their insights on ethical frameworks throughout the entire work. PopAI partners have also liaised with INTERPOL and UNICRI and reviewed their ethics toolkit for AI in law enforcement.

Finally, we relied on the input of PopAI, ALIGNER and STARLIGHT LEAs partners who provided their questions regarding AI and ethics. The questions were organized into five clusters (refer to Appendix A) and served as the foundation for guiding the content and organization of this report. The questions will also serve as a basis for the development of educational materials which will be incorporated into the ethics toolkit (D2.5).

2 What is ethics and AI ethics?

Ethics is a branch of philosophy that deals with moral principles and values (Rességuier & Rodrigues, 2020). It is concerned with what is right and wrong, good and bad, and fair and unfair in human behaviour and decision-making. Ethics is a fundamental aspect of human society and it plays an important role in shaping our interactions with others, guiding us to make morally right decisions and holding us accountable for our actions (Driver, 2006). Ethics can be divided into several branches, such as normative ethics, which deals with determining broad moral principles, and applied ethics, which deals with the practical application of these principles to specific situations and fields, such as business, healthcare, and technology (Manners, 2008). While normative ethics define what is acceptable and unacceptable, desirable or undesirable, and our conception of when things go well and when things should go differently, applied ethics provides a framework for evaluating the ethical implications of actions and decisions and for determining the responsibilities and obligations of individuals and organizations in specific sectors (Cahn & Markie, 2019).

Building a culture of ethics involves the introduction of different mechanisms that are distinct but related to one another. These mechanisms include lists of principles, ethics guidelines, ethics recommendations and codes of ethics.

2.1 Ethics principles

Ethics principles are the fundamental moral values that are used to guide ethical decision-making and behaviour. These principles are often derived from various ethical theories and can vary depending on the context or the field of study. However, there are some general principles that are commonly accepted across different fields and cultures.

Some common ethics principles include (Singer, 2011):

- Autonomy: respect for individuals' freedom of choice and ability to make decisions for themselves.
- Non-maleficence: the principle of "do no harm," which holds that individuals and organizations have a moral obligation to avoid causing harm to others.
- Beneficence: the principle of "doing good," which holds that individuals and organizations have a moral obligation to promote the well-being of others.
- Justice: the principle of fairness and equity, which holds that individuals and organizations have a moral obligation to treat all people fairly and equitably.
- Transparency: the principle of being open and honest, which holds that individuals and organizations have a moral obligation to be transparent in their actions and decisions.

Depending on the field and context, other principles such as confidentiality, responsibility, integrity, can be included. It's important to note that these principles are not mutually exclusive, and that ethical decision-making often involves balancing different principles (Brenkert, 2009; Weinstock, 2013). Additionally, different situations may require different priorities of ethical principles (Van den Hoven et al., 2012).

2.2 Ethics guidelines

Ethics principles and ethics guidelines are related but distinct concepts. Ethics principles are broad, fundamental beliefs or values that provide a general framework for decision-making and behaviour (AI HLEG, 2019). They are often abstract and general in nature and are not specific to a particular profession or context. Examples of ethical principles include honesty, fairness, and respect for others. Ethics guidelines, on the other hand, are more specific and concrete, and are often developed by professional organizations to provide practical guidance for members of a particular profession or industry (e.g. AI HLEG, 2019). They are usually based on the broader ethics principles and provide more specific rules and recommendations for how to behave in a particular context or situation. For example, a professional organization for law enforcement might develop ethics guidelines that specify how officers should interact with members of the community or handle sensitive information. In short, ethics principles provide the foundation for ethical decision-making and behaviour, while guidelines are more specific and concrete applications of those principles.

One example of an ethical guideline for law enforcement is "protect and serve."(Council of Europe, 2001). This principle states that the primary duty of law enforcement officers is to protect and serve the community, and to do so in a fair and impartial manner, without discrimination or bias. This principle is based on the idea that law enforcement officers have a special responsibility to uphold the law and protect the rights and safety of all members of the community, regardless of their race, gender, religion, or any other characteristic.

2.3 Ethics recommendations

Ethics guidelines and ethics recommendations are similar concepts, but they can have slightly different meanings depending on the context. While ethics guidelines are specific rules or principles that are intended to provide practical guidance for ethical decision-making and behaviour within a particular profession or industry. They are often developed by professional organizations and are intended to help members of the profession or industry navigate complex ethical issues and dilemmas. Guidelines are usually binding, enforceable and organizations have a responsibility to follow them. Ethics recommendations, on the other hand, are suggestions or advice for how to approach a particular ethical issue or problem. They may be less formal than guidelines and may not be binding (e.g. Rubio et al., 2020). Recommendations can be made by professional organizations, individuals, or a group of experts. They are often intended to provide guidance and direction for decision-making, but they may not be mandatory to follow. In short, guidelines are more formal and binding whereas recommendations are less formal.

2.4 Ethics codes

Ethics codes, also known as codes of ethics, are formal statements that outline the values and principles that an organization or profession adheres to (Council of Europe, 2001a). They are designed to provide guidance for decision-making and behavior in accordance with the organization's or profession's mission, values and goals. Ethics codes can take various forms, such as a written document, a set of guidelines, or a set of rules. They can be specific to an organization or profession, or they can be more general and apply to a broader group of individuals or organizations. Ethics codes can address a wide range of topics, such as integrity, honesty, fairness, confidentiality, responsibility, and respect for others. They can also include specific procedures for reporting and addressing ethical violations. Ethics codes are intended to serve as a benchmark for ethical behaviour and decision-making. They are meant to be used as a reference when facing ethical dilemmas, to help individuals and organizations act in a morally responsible way, and to ensure that their actions align with their values and mission. It's important to note that, while having an ethics code is important, it is not enough. It's also important to have a culture of ethics, where the code is understood, supported, and implemented in day-to-day decision-making, and where there are mechanisms in place to ensure compliance and to address any violations (Sinclair, 1993).

2.5 What do ethics mechanisms aim for?

Ethics principles, guidelines, recommendations, and codes provide ethical guidance for a variety of activities. They establish principles, beliefs, and standards of ethical and professional behaviour as well as specify expectations, restrictions, or prohibitions on conduct and activities. Generally, ethics mechanisms strive to encourage individuals' reflection on behaving with benevolence and without causing harm to others (Driver, 2006).

Specifically, ethics principles, guidelines, recommendations and codes strive to (United Nations, 1979; Zardiashvvili et al., 2019):

- define the values and principles that the organisation and profession uphold and build a common ground of shared meanings and understandings;
- raise concerns about the ideals upheld by an organisation as a whole and how they should be applied and play an important role for training and trainers;
- promote ethical behaviour;
- raise awareness on specific challenges;
- facilitate the detection and resolution of issues;
- indicate the circumstances in which potential acts and behaviours conflict with the duties ;
- increase trust within the organisation as well as between the organisation and the public;
- facilitate supervision;
- make people more accountable;
- give guidance on what should be done in specific situations;

Ethics mechanisms can be very helpful in fostering an organizational culture of integrity, informing all personnel of their ethical obligations, and in leveraging peer pressure to advance moral values and fortify integrity. A culture of ethics increases the likelihood that issues, harms and risks will be more easily detected and thoroughly understood, more thoroughly analyzed, and more easily solved (Northern Ireland Policing Board, 2008).

2.1 AI Ethics

AI ethics is a branch of ethics that defines a set of values, principles and techniques to guide moral conduct in the development and use of AI (Coeckelbergh, 2020). AI ethics considers the ethical issues that arise from the development and use of AI systems and provides guidance to assure that AI systems are developed and used without harm and in a way that maximise AI benefits for all.

Some of the key issues that AI ethics addresses include (Leslie, 2019):

- Ensuring that AI systems are not biased and do not perpetuate discrimination;
- Ensuring that AI systems are transparent and explainable, so that their decisions can be understood and justified;
- Ensuring that AI systems are accountable and that there is human oversight of their use;
- Ensuring that AI systems are safe and do not cause harm;
- Ensuring that AI systems respect and protect civil rights, privacy and data rights.

With AI systems becoming increasingly part of our society and with society being increasingly aware of its risks, the field of AI ethics is booming to guide ethical development and use of AI (Jobin et al., 2019)[1].

2.2 Why do we need AI ethics in Law Enforcement?

To see why AI ethics, need to be taken seriously by LEAs, it is important to understand how AI can be used in law enforcement and what are the risks. Artificial Intelligence (AI) in law enforcement AI can be used in various ways (see D2.1 for more details):

- **Data analytics and prediction:** AI technologies can analyze large datasets, identify patterns and extract insights from the data fed into the systems. In law enforcement, data analytics powered by AI are used to predict crime spots, offenders, perpetrators' identities or victims of crime.
- **Machine vision & recognition:** AI technologies can identify or understand images and videos. Facial recognition technologies and licence plate readers are examples of this class of tools that can trace and track individuals and objects.
- **Natural language processing** is a family of techniques that make human language decipherable by a software. NLP techniques can analyze and generate language data to extract insights. AI powered chatbots using NLP can be used to interact with the public and provide information about crime prevention and reporting.
- **Process automation**: AI technologies have the capacity to automatically carry out standardized activities under well-defined conditions. For example, autonomous patrol vehicles are self-driving vehicles equipped with AI that can patrol areas. AI can also be used to automatically detect cybercrime such as hacking.

These functions might create serious harms such as (see D2.3, D2.2 and D3.1):

[1] See Algorithm Watch, "AI Ethics Guidelines Global Inventory"

- **Breach of privacy**: AI systems are grounded in the processing of data; therefore they might involve the use of personal data. Data might be used without the consent of the data subject and might be processed in ways that do not guarantee anonymisation, thus revealing personal information (see D2.2). Citizens' privacy can be undermined by decisions taken in the development and design of the AI system, as well as by the use of the AI systems to track individuals. Targeting or profiling individuals can infringe on the ability of citizens to lead a private life in which they can actively control the effect of technology on them.
- **Bias and discrimination:** Data-driven AI applications can reproduce, perpetuate, and magnify patterns of marginalization, inequality, and prejudice that already exist in societies since they gain insights from existing and past structures and dynamics. AI designers' and developers' preconceptions and biases can affect AI systems through their many features and analytic structures. Pre-existing biases can affect AI models from the definition of its goal, data collection, building and training of the model and the user interaction with the model. For example, the COMPAS software, which measures the risk of a person re-offending have been found to be biased toward African American individuals (Angwin et al., 2016) . Biases in algorithmic systems can lead to discrimination. Technical bias that involves the underrepresentation of protected attributes (e.g. age) can also be introduced in the AI model during its development. Bias can also emerge when users interact with the systems, or the data employed to build it.
- **Non-transparent, unreliable, or unexplainable outcome:** AI systems might rely on correlation and regression models that are not easily interpretable by the person who needs to take a decision causing unreliable, unsafe and biased outcomes.

More specifically, AI functionalities are being increasingly used by law enforcement agents for crime prevention and crime investigation, in the management of migration, the administration of justice, cyber operations, as well as training of LEAs (see D3.1). Each of these area raises important and context-specific ethical concerns that are outlined below.

2.2.1 Crime prevention and investigation

There are several issues associated with using AI tools that assess areas and individuals to make predictions, profiles and assess the risk of (re)offending including (Fair Trials, 2021; see also D2.3, D3.1):

- **Bias**: Based on the data used to train them, AI systems can perpetuate existing discriminatory practices in policing and even amplify them at a larger scale, resulting in broader unfair or discriminatory outcomes that target people based on their ethnicity, race, socio-economic status and environment.
- **Lack of transparency**: It can be difficult to understand the data variables included in the predictive or risk assessment model, how weights were assigned to the variables, how AI

systems assigned a risk score and its accuracy, making it challenging to hold users accountable and to identify and correct errors.

- **Privacy concerns**: AI systems that are used to assess risk and prevent crime raise concerns about the collection, storage, and use of personal data[2].
- **Misuse**: if they are not properly regulated, monitored and used AI-powered tools to prevent crime and assess risk can be used to violate civil liberties and human rights such as the presumption of innocence, the right to have a fair trial, or the right to privacy[3].
- **Reliance on AI**: Overreliance on AI systems may lead to a loss of critical thinking and decision-making skills among law enforcement personnel when making decisions about risk and crime prevention (Park et al., 2021).
- **No rehabilitation**: Emotion is crucial to intelligent behaviour and decision-making (Bechara et al., 2000). AI systems' output rely on past information and have no emotional intelligence affecting their output (Schuller & Schuller, 2018). While previous behaviour may provide insight into the present and future actions, it does not consider the idea of an opportunity for rehabilitation, which has the consequence of reinforcing negative beliefs and punishing people who might have already paid their debt.

- **Self-fulfilling prophecy**: data used for predictive policing which have been demonstrated to reflect patrolling priorities in disadvantaged areas rather than criminal activities, risking leading to patrolling decisions that contribute to self-fulfilling prophecies : if police are dispatched to a specific area, it will follow that more crime will appear there (King & Mertens, 2023).

2.2.2 Management of Migration

There are several ethical issues associated with using artificial intelligence (AI) in the management of migration and borders by law enforcement, including (see D2.2, D2.3, D3.1):

- **Design**: Migrants are frequently depicted as a security danger during the design and testing of algorithmic technologies rather than as humans with rights and liberties (Bircan & Korkmaz, 2021). This affects negatively the technology lifecycle from its design to its use.
- **Lack of transparency**: AI systems can be opaque and difficult to understand, making it challenging to hold law enforcement accountable for their actions when it comes to decisions on migrants' and refugees' lives (Beduschi, 2022).
- **Privacy concerns**: The use of AI-powered surveillance and tracking technologies raises concerns about privacy and civil liberties, particularly for migrants and refugees who are systemically put in a vulnerable position (Molnar, 2020).

[2] Art.4 GDPR: 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[3] https://www.europarl.europa.eu/charter/pdf/text_en.pdf

- **Lack of oversight**: The use of AI in migration management may be subject to little or no oversight (Molnar, 2021), which can lead to abuse and misuse of the technology by law enforcement.
- **Inadequate data**: AI systems require large amounts of data to function, but data about migration is often limited and inadequate (Singleton, 2016), which can lead to inaccurate or incomplete information being used to make decisions on people who are in a vulnerable situation.
- **Lack of understanding**: Many law enforcement agencies may not have the technical expertise or knowledge to understand and effectively use AI, which can lead to ineffective or even harmful outcomes for migrant people and refugees.

2.2.3 Administration of Justice

There are several issues with using artificial intelligence in the administration of justice, including (see D2.2, D2.3, D3.1):

- **Bias**: AI systems can be trained on biased data, which can lead to biased decisions. This can disproportionately affect marginalized groups and lead to unfair outcomes.
- **Lack of transparency**: AI systems can be complex and difficult to understand, making it difficult to explain how a decision was reached. This can make it difficult to hold decision-makers accountable for the outcomes of the system.
- **Error-prone**: AI systems are not perfect and can make mistakes, which can lead to wrongful convictions or acquittals (Fair Trials, 2022).
- **Lack of human oversight**: AI systems can be used to automate decision-making, which can lead to a lack of human oversight and accountability.
- **Privacy**: AI systems can collect, process and store large amounts of personal data, which can raise privacy concerns.

2.2.4 Cyberoperations

There are several issues to consider when using AI for law enforcement to combat cybercrime, and conduct operations in the cyberspace including (see D2.2, D2.3, D3.1):

- **False positives:** AI tools for scanning online (eg. child sexual abuse) materials can suffer from low accuracy and high rates of false alarms and lead to target innocent individuals (EDRi, 2022).
- **Stability of cyberspace:** there is a trade-off between AI for benevolent and malevolent use in the cyberspace. On the one hand, developments in AI give attackers more options for targeting and new ways to deliver attacks, enabling them to utilize more sophisticated and thorough operations. On the other AI enhances response and defense mechanisms (Taddeo, 2019).
- **Privacy:** AI can verify users by profiling them based on the unique way they use their mouse. AI can make systems more resistant to attacks when more data are made available. This poses under threat user privacy, it expose them to extra risks if confidentiality is breached and might lead to mass surveillance (Taddeo, 2013).

2.2.5 Training of LEAs

There are several issues to consider when using AI applications such as virtual reality (VR) for training law enforcement, including (see D2.2, D2.3, D3.1):

- **Safety**: VR training can be intense and may cause physical or psychological harm to trainees, especially if it is not properly designed and supervised (Al-Jarani, 2019).
- **Reliability**: VR simulations may not fully replicate real-world situations, which can lead to a lack of realism in training and a lack of transferability of skills to actual field scenarios.
- **Bias**: VR training simulations may be based on biased assumptions and stereotypes, which can lead to biased behaviour among trainees.
- **Accountability and abuse**: training law enforcement officers in a VR environment raises issues of accountability. There might be concerns about the use of virtual force, as the actions of officers in virtual environments could be interpreted as excessive force in real-life situations
- **Technical issues**: VR training systems may have technical issues such as latency, resolution and stability, which can affect the training experience and the results (Anthes et al., 2016).
- **Limited interpretability of the performance:** VR based training may be limited in their interpretability, making it difficult to understand how the trainee is performing. Evaluating the performance of an officer in a virtual environment can be challenging, as it may not accurately reflect the capabilities in real-life situations. This can make it difficult to assess an officer's level of competency.

Therefore, It is important to carefully consider the challenges that AI might pose in different areas of law enforcement and take legal and ethical steps to mitigate the risk and reduce the harm. Creating a culture of responsible and ethical AI is key to have an ethical development and use of AI. The taxonomy presented in the next section aims to aid law enforcement officers and relevant stakeholders to steer the field of AI ethics into shaping a culture of AI ethics in their environment.

3 The Ethics taxonomy

This section presents the PopAI ethics taxonomy. This taxonomy wants to help LEAs, researchers, policymakers and citizens interested in AI ethics and Law enforcement to:

- navigate the field of AI ethics;
- create a culture of AI ethics;
- raise awareness on AI ethics in law enforcement;
- take decisions on policies, organisational strategies, research priorities and training.

3.1 Methodology

In this work, we combined a semi-systematic literature assessment of AI ethics guidelines with both unstructured and structured thematic analysis methods. A group of five researchers used these methods to independently identify themes, trends, and gaps in a body of literature, jointly organize the themes into a common analytical framework, and synthesize principles into the new PopAI ethics guidelines for LEAs use of AI.

Our review focused on principles, guidelines and scientific work on **LEA ethics, AI ethics** or a **combination of AI and LEA ethics.**

To create the taxonomy and the popAI ethics guidelines for law enforcement, we divided the research process into four phases.

1. **Phase 1: Document set creation**. Each researcher was instructed to provide a list of documents -including guidelines, codes of conduct, principles, tools and scientific research-related to ethics for Law Enforcement, Artificial Intelligence, or both.

2. **Phase 2: Taxonomy creation**. For each document, researchers had to indicate:
   - **general features** that included the title, year, authors, and link
   - **the actor** that could either be:
     - academia or research institute: pieces of work produced in the academic sector or research sector
     - professional association: pieces of work produced by companies and organisations with specific competencies in a field (e.g., IEE)
     - government: pieces of work produced by policymaking bodies (e.g., EC, Council of Europe).
     - intergovernmental organisations: pieces of work produced by intergovernmental organisations (e.g. UN)
     - civil society: pieces of work produced by civil society organisations (e.g. Amnesty)
     - private sector: pieces of work produced by any private player (e.g. Microsoft).
   - whether the document contained a:

    - a high-level list of principles or
    - information on how to operationalise the principles.
- the **type of document**:
    - policy: ethics documents related to policy work;
    - recommendations: ethics documents with recommendations;
    - guidelines: ethics documents that provided guidelines;
    - principles: ethics documents with broad principles;
    - tool: ethics documents that provided practical tools (e.g. algorithmic, self-assessment lists) to address ethics.
    - code of ethics
- the **approach** of the document that could wither be:
    - **Horizontal**: when the document was applying to all applications of AI across all sectors;
    - **Vertical:** applying to only a specific application of AI or a specific industry.
- the field: AI, LEAs or AI/LEAs.
- the **functionality/ies** mentioned.
- Lessons for developing EU ethics principles for Law Enforcement
- General aspects of interest.

3. **Phase 3**: **Comparison and triangulation.** Researchers discussed the documents according to the categories explained above. Researchers triangulated common topics, trends and gaps and discussed potential approaches to structure the PopAI guidelines. At this stage, researchers decided to create a list of guidelines that merged LEAs ethics principles and AI ethics principles.

4. **Phase 4: popAI guidelines creation**. Researchers compared all principles identified that were similar or repeated through the documents. Analyzing and synthesizing the most common topics enabled the development of key guidelines for AI in law enforcement coupled with recommendations for their implementation.

3.2 Outcome of Phase 2: Taxonomy creation

The result of this phase is attached to Appendix B which contains the popAI Ethics taxonomy. The Ethics Taxonomy is meant to be updated with relevant documents until the end of the popAI project - September 2023.

It is important to note that the taxonomy is not intended to serve as a comprehensive list, but rather as a categorization framework for better understanding and organizing the field of AI and LEA ethics with references to specific types of documents, principles and functionalities. While efforts have been made to include a broad range of works on ethics, it is acknowledged that there may be additional works that have not been included. Therefore, taxonomy should not be viewed as exhaustive or definitive, but rather as a living tool for analysis and discussion.

3.3 Outcome phase 3: Comparison and triangulation

This section describes the comparative and triangulation study phase. To maintain conciseness, we will highlight the key Law Enforcement Agencies (LEAs), Artificial Intelligence (AI), and LEAs-AI ethics documents included in the taxonomy. We will also outline the prominent LEAs and AI principles that will inform the popAI ethics guidelines outlined in Section 4.

3.3.1 Ethics in Law Enforcement

It is important to note that each country has different codes of conduct and laws, and that there may be regional variations within the country as well. Also, not all countries have a specific Code of Police Ethics, but they have laws, regulations and policies that govern the behaviour and conduct of police officers. The main codes of police ethics include the European Code of Police Ethics, the Police Officer European Charter, the United Nations Code of Conduct for Law Enforcement Officials, The United Nations Basic Principles on the Use of Force and Firearms by Law Officials and the International Association of Chiefs of Police (IACP) Code of ethics which are discussed in the following sections.

3.3.1.1 The European Code of Police Ethics

The European Code of Police Ethics was adopted by the Committee of Ministers of the Council of Europe in 2001 (Council of Europe, 2001). It sets out basic principles and standards for police conduct in Europe with particular emphasis on human rights, training, the relationship with the public and the accountability of officers. The code is divided into seven main sections.

Section one defines the **objectives of the police** that are: maintaining public tranquility and law and order, to protect and respect the individual’s fundamental rights and freedoms as enshrined, in particular, in the European Convention on Human Rights; to prevent and combat crime; to detect crime; to provide assistance and service functions to the public. It emphasizes the trend in Europe to integrate the police more fully into civil society through community policing and giving the police the status of a public service body.

Section two defines **the legal framework for the police** as an institution and its actions according to the rule of law. It states that the police are a public body established by law and must always conduct operations in accordance with national and international standards. The legislation guiding the police should be accessible to the public and clearly formulated. Police personnel should be subject to the same legislation as ordinary citizens, with exceptions only justified for the proper performance of police duties in a democratic society. The principles outlined in this section are based on the European Convention on Human Rights and its accompanying case law.

Section three discusses the **role of the police in the criminal justice system** and the importance of a separation of the role of the police from the other components of the criminal justice system. It emphasizes the need for the police to respect the independence and impartiality of judges and not to hinder their decisions. It also states that the police should not have judicial functions, but in certain situations, they may be entrusted with them, but these decisions can be challenged before a court. The text also mentions the importance of functional and appropriate cooperation between the police and the public prosecution, where the police should receive clear instructions and keep the superior crime investigation authorities informed.

Section four discusses the **organisation of the police in a democratic society** governed by the rule of law. The main themes include the importance of earning public respect, promoting professionalism, transparency, and accountability within the police force. It also stresses the importance of the police being under civilian authority, being easily recognizable, having operational independence, and having a clear chain of command where superiors are held accountable for the actions of subordinates. It also notes that the the police structure should promote cooperation with other agencies and good public relations which include being ready to give objective information on their activities to the public, without disclosing confidential information. It discusses the **recruitment and retention of police personnel.** It states that candidates should be selected based on their personal qualifications and experience, which should be appropriate for the objectives of the police. It also notes that police personnel should be able to demonstrate qualities such as sound judgment, fairness, and communication skills, and should possess a good understanding of social and community issues. Additionally, the passage states that persons who have been convicted of serious crimes should be disqualified from police work, and that recruitment procedures should be based on objective and non-discriminatory grounds with the goal of recruiting a diverse group of individuals. The passage also mentions that **police training** should be based on the values of democracy, the rule of law, and the protection of human rights and that training shall be as open as possible towards society.

Section five provides **guidelines for police intervention** with general as well as more specific guidelines. General principles include respect everyone's right to life, restriction on the use of force, impartiality and non-discrimination, respect to privacy, restriction on the use of personal data, opposing corruption. Guidelines for specific situations include principles for police investigation and for the deprivation of liberty. The Code suggests that police investigations shall, as a minimum, be based upon reasonable suspicion ofan actual or possible offence or crime; that the police must follow the principles that everyone charged with a criminal offence should be considered innocent until found guilty by a court. and that everyone charged with a criminal offence has certain rights, in particular, the right to be informed promptly of the accusation against them and to prepare the defence; that police investigations shall he objective and fair. They shall be sensitive and adaptable towards the special needs of persons, such as children, juveniles, women, minorities, including ethnic minorities. and vulnerable persons; that police shall provide support and information to the victims of crime without discrimination.

Section five summarizes the key **accountability principles** of the police which are particularly relevant when it comes to using AI tools. The Code states that police shall be accountable to the state, the citizens and their representatives and that accountability mechanisms, based on communication and mutual understanding between the public and the police, shall be promoted. They shall be subject to efficient external control.

Finally, the last section discusses **research and international cooperation** involving the police. It encourages research on the police and international cooperation on police ethics and human rights aspects.

3.3.1.2 The Police Officer European Charter

The **Police Officer European Charter** was developed by the European Council of Police Trade Unions in 1992. It lays out principles for the organization and function of police services, with the goal of ensuring that the police are a public service that serves the community and guarantees citizens' rights and freedoms. It revolves around three key points: the **organisational aspects, the relationship with the community and the personal and professional status of police officers**. The charter emphasizes the importance of preventing illegitimate political interference in police activities, demilitarizing police services, protecting citizens' rights, being transparent in actions and activities, and limiting police action to combat criminal activity. When laying out the guidelines for relationships between police and the community, the Charter emphasizes the importance of treating citizens correctly, protecting citizens and their rights, and respecting the honour and dignity of individuals. While principles for the personal and professional status of police officers include the right to form trade unions, receive proper training, have just salaries, have appropriate working conditions, and have legal defense and social protection.

3.3.1.3 The United Nations Code of Conduct for Law Enforcement Official

The **United Nations Code of Conduct for Law Enforcement Official** (United Nations, 1979) was adopted by the General Assembly of the United Nations in 1979 and sets out principles and standards for the behavior of law enforcement officials around the world. It is divided into 8 articles detailing the responsibilities and conduct of law enforcement officials. Law enforcement officials must at all times fulfill their duty to serve and protect the community (Article 1), while also respecting and protecting human dignity and upholding human rights (Article 2). They are only allowed to use force when necessary and to the extent required for their duty (Article 3), and must keep confidential information they possess confidential unless necessary for their duty or justice (Article 4). They are prohibited from inflicting, instigating, or tolerating any act of torture or cruel, inhuman, or degrading treatment or punishment, and may not use superior orders or exceptional circumstances as justification (Article 5). They must ensure the full protection of the health of persons in their custody

(Article 6). They must oppose and do not commit corruption (Article 7) and respect the law of the code (Article 8)

3.3.1.4 The United Nations Basic Principles on the Use of Force and Firearms by Law Enforcement Officials

This set of principles, adopted by the General Assembly of the United Nations in 1990 (United Nations, 1990), provides guidelines for the use of force and firearms by law enforcement officials. It states that governments and law enforcement agencies should establish rules and regulations on the use of force and firearms, while keeping ethical considerations in mind. It also emphasizes the development and use of non-lethal incapacitating weapons and self-defense equipment to decrease the need for lethal force. Law enforcement officials are instructed to use non-violent means before resorting to force and firearms, and to use force and firearms only when necessary, with restraint and in proportion to the situation. The passage also states that any injury or death caused by the use of force and firearms must be reported and that the arbitrary or abusive use of force and firearms must be punished as a criminal offense. Additionally, it is mentioned that exceptional circumstances should not be used to justify any departure from these guidelines.

It states that rules and regulations for the use of firearms by law enforcement officials should include guidelines for when firearms can be carried and used, the types of firearms and ammunition permitted regulations for controlling, storing, and issuing firearms, and a system for reporting when firearms are used in the performance of duty. This document also stresses the importance of training which should include issues of police ethics and human rights, especially in the investigative process, to alternatives to the use of force and firearms, including the peaceful settlement of conflicts, the understanding of crowd behaviour, and the methods of persuasion, negotiation and mediation, as well as to technical means, with a view to limiting the use of force and firearms.

3.3.1.5 The International Association of Chiefs of Police (IACP) Code of Ethics and Standards of conduct

The International Association of Chiefs of Police (IACP) code of ethics was adopted by IACP in 1957 (IACP, 2019), sets out principles and standards for the conduct of police officers in the United States and internationally. It stresses the importance of serving the community, avoid prejudice and discrimination, public trust, and responsibility.

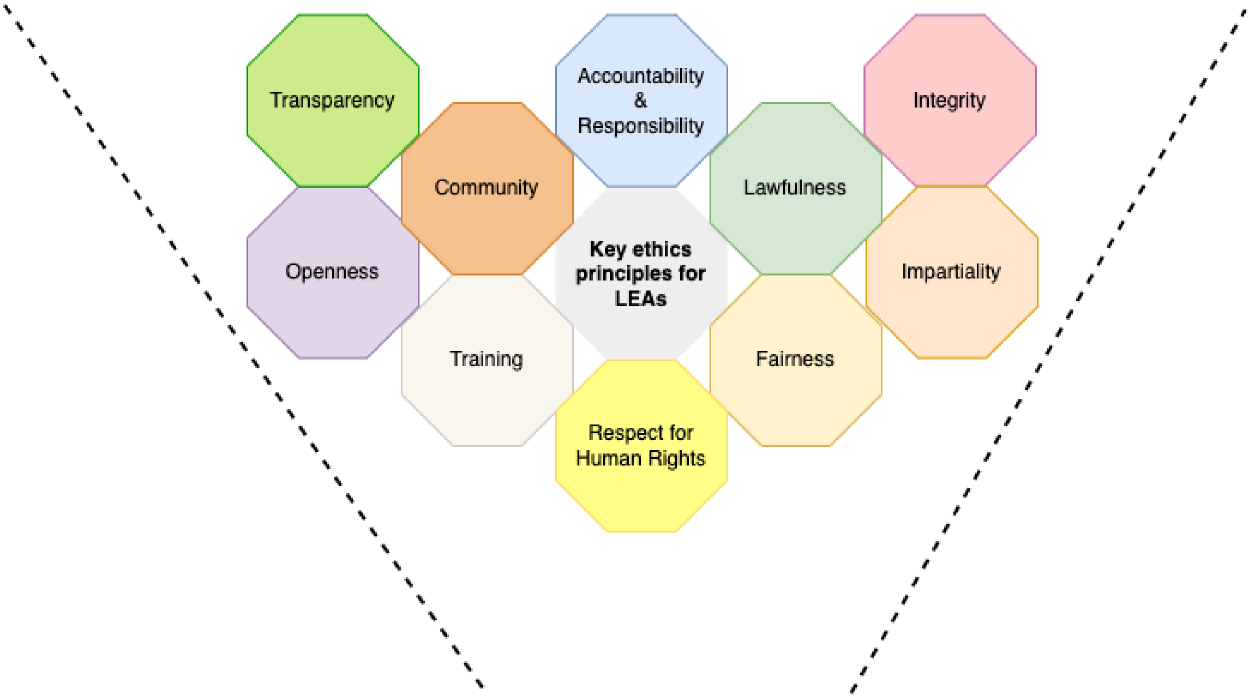
The IACP Standards of conduct provide guidance on LEAs standard of conducts which include: adherence to law, accountability and responsibility, guidance on public statements and appearance and political activities.

3.3.1.6 Key LEA principles

Despite the substantial variations in the principles regulating the professional behavior of law enforcement across different nations and organizations, the reviewed documents (as listed in the taxonomy) encompass the following values:

- **Lawfulness**: operations must always be conducted in accordance with the national law and international standards accepted by the country (United Nations, 1979).
- **Accountability and responsibility**: LEA personnel is accountable to the State, citizens and their representatives for their actions and omissions (Council of Europe, 2001). In this regard, it is important to provide accountability mechanisms, based on the mutual understanding between the public and the police.
- **Openness**: LEA personnel should be aware of social, cultural and community issues and they should demonstrate sound judgment, an open attitude and maturity, fairness, and communication skills (Council of Europe, 2001).
- **Integrity**: LEAs should "do the right thing" consistently and reliably (EUROPOL, 2019) and prevent and combat police corruption (United Nations, 1979).
- **Fairness**: the police act with fairness when they show full respect for the positions and rights of each individual that are subject to their police duties (Council of Europe, 2001; United Nations, 1979). Fairness should apply to all aspects of police work, but it is particularly emphasised with regard to the public (Council of Europe, 2001).
- **Transparency**: the police should be as transparent as possible towards the public (United Nations, 1979).
- **Impartiality**: police act with integrity and with a view to avoiding taking sides (EUROPOL, 2019).
- **Respect for human rights:** the police conduct must protect human dignity, respect the right to non-discrimination, the right to privacy and maintain and uphold the human rights of all persons (Council of Europe, 2001; United Nations, 1979, 1990).
- **Community**: the police is an integrated part of society and as such, it shall foster positive relationships with other public bodies and the civil society (Council of Europe, 2001). In this regard, the police should be considered a resource and a service to the civil society instead of a force imposed and is called to give objective information on their activities to the public and collaboratively solve problems with citizens.
- **Training**: governments and law enforcement agencies shall ensure that all law enforcement officials are provided with training. Police must be trained with special attention to issues of ethics and human rights the fundamental values of democracy, rule of law and protection of human rights. The personnel must be trained in an environment, which is as close as possible to social realities (Council of Europe, 2001).

The extraction of values has resulted in the identification of six fundamental guidelines (LEA-a/b/c/d/e/f) that outline the ethical function and responsibilities of Law Enforcement agencies within a democratic society. These guidelines will be thoroughly examined and related to to AI principles in Section 4.



(LEA-a) Ensuring/promoting public safety, including preventing crime
(LEA-b) Exercising restraint when intervening, especially when intervening forcefully
(LEA-c) Treating individuals fairly and impartially
(LEA-d) Integrity and respect for the rule of law
(LEA-e) Protecting and promoting the rights of individual citizens
(LEA-f) Serving the community and building relationships

Figure 1: Key ethics principles and guidelines for law enforcement

3.3.2 Ethics in AI

AI ethics guidelines began to be developed from the 2010s as it became clear that there were potential ethical implications and consequences of AI use. These guidelines aim to ensure that AI is developed and used in a way that is safe, fair, and beneficial for society as a whole. Additionally, some of the guidelines are developed to prevent AI from being used to discriminate or perpetuate bias. The development of AI ethics guidelines is an ongoing process, as new technologies and use cases continue to emerge. In recent years, there has been a significant increase in the number of organizations and groups that have developed or are in the process of developing AI ethics guidelines. This includes government agencies, industry groups, and academic institutions. The most notable example in the EU is the Ethics Guidelines for Trustworthy AI.

3.3.2.1 EU Ethics Guidelines for Trustworthy AI

The European Commission established a group of experts - the High Level Expert Group on Artificial Intelligence- to provide advice on its artificial intelligence strategy (AI HLEG, 2019). In 2019, the AI HLEG produced the Ethics guidelines for trustworthy AI which identify key ethical principles and the values that must be respected in the development, deployment and use of AI systems. These guidelines consider AI a mean to increase human flourishing, enhancing well-being and the common good and are based on fundamental rights and identify three components that should be met through the system's lifecycle:

- AI systems should be **lawful** so they should comply with European, national and international laws and regulations
- AI systems should be **ethical** so they should adhere to ethical principles and values
- AI systems should be **robust** both from a social and technical perspective, to prevent the harm it might cause.

To offer guidance on the implementation of these three key principles, the AI HLEG developed 7 key requirements that need to be met by developers, deployers, decision-makers and end-users to achieve trustworthy AI:

- **Human Agency and oversight** including fundamental rights, human agency and human oversight
- **Technical robustness and safety** Including resilience to attack and security, fall back plan and general safety, accuracy, reliability, and reproducibility
- **Privacy and data governance** Including respect for privacy, quality and integrity of data, and access to data
- **Transparen**cy including traceability, explainability and communication
- **Diversity, non-discrimination and fairness** including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation
- **Societal and environmental wellbeing** including sustainability and environmental friendliness, social impact, society and democracy
- **Accountability** including auditability, minimisation and reporting of negative impact, trade-offs and redress.

The AI HLEG produced also an Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (AI HLEG, 2020).

3.3.2.2 Key AI principles

The analyzed AI ethics documents, as specified in the taxonomy, encompass the following principles:

- **Respect for human rights** in AI refers to the principle of ensuring that artificial intelligence systems are designed and operated in a manner that respects fundamental human rights such as human dignity, privacy, freedom of expression, and non-discrimination (Leslie, 2019). This includes ensuring that AI systems are not used to violate human rights, such as through surveillance or censorship, and mitigating risks such as discrimination or bias. Respect for human rights in AI also involves transparency and accountability in the development and deployment of AI systems, as well as active efforts to identify and address potential human rights violations. The goal of respect for human rights in AI is to ensure that AI is used in a manner that respects and protects the basic rights and dignity of all individuals (UNESCO, 2021).

- **Human Agency and oversight** in AI refers to the role that humans play in controlling and directing the actions and decisions of AI systems (AI HLEG, 2019). This includes setting ethical guidelines, monitoring AI systems for bias and error, and ensuring that AI systems are aligned with human values (Future of life Institute, 2017) and goals throughout all the lifecycle. The goal of human agency and oversight is to ensure that AI systems are safe, transparent, and accountable, and to ensure that AI is used for the benefit of humanity.
- **Accountability** in AI refers to the responsibility of AI system creators, operators, and decision-makers for the actions and outcomes produced by AI (IEEE, 2019). This includes ensuring that AI systems are transparent, ethical, and aligned with societal values and that they do not cause harm. It also involves having mechanisms in place to investigate and address any negative consequences that may result from AI systems, and being able to explain and justify the decisions made by AI systems (OECD, 2019). The goal of accountability is to ensure that AI is used in a responsible and trustworthy manner.
- **Explainability** in AI refers to the degree to which the decision-making processes of an artificial intelligence system can be understood and interpreted by humans (OECD, 2019). It is the extent to which the outputs of a model can be accounted for and justified based on the input data and the methods used to generate the results. The goal is to provide transparency and understanding of how AI systems reach their decisions, to ensure accountability, reliability, and fairness.
- **Socio-technical robustness** in AI refers to the consideration of both technical and social factors when designing and deploying AI systems (AI HLEG, 2019). This includes not only ensuring the technical robustness and safety of the AI system, but also ensuring that the system operates in an ethical, transparent, and socially responsible manner. This involves considering the social impacts of AI, such as potential biases and unintended consequences, and designing AI systems that are aligned with human values and goals. The goal of socio-technical robustness in AI is to create AI systems that are not only technically sound, but also socially responsible and beneficial for society as a whole.
- **Safety or security** in AI refers to the measures taken to ensure that artificial intelligence systems operate in a manner that minimizes risk of harm to humans and the environment (Toreini et al., 2020). This includes avoiding unintended consequences, such as accidents or harm to human life, as well as mitigating ethical risks, such as bias or discrimination. Safety in AI also involves ensuring that AI systems are secure against malicious attacks and data breaches. The goal of safety in AI is to design and deploy AI systems that are trustworthy, reliable, and safe for human use.

- **Transparency** in AI refers to the quality of AI systems being easily understood and interpretable by human users (AI HLEG, 2019). This includes being able to explain how the AI system made its decisions, the data it was trained on, and any biases that may exist within the system. The goal of transparency is to increase trust in AI and ensure ethical and fair outcomes.
- **Societal and environmental well-being** in AI refers to the impact that AI technology has on individuals and the world as a whole (Floridi et al., 2018). This includes considering the ethical, social, and environmental consequences of AI systems and ensuring that they align with societal values and contribute positively to the well-being of individuals and the environment. This can involve designing AI systems that are inclusive, reduce inequalities, minimize harm, and promote sustainability. It implies that AI designers and users must be sensitive to the wide range of cultural norms and values.
- **Accuracy** in AI refers to the ability of an AI system to produce the desired results, achieve its goals and objectives, and deliver value to its users. The goal of accuracy is to ensure that AI systems are able to perform well and provide tangible benefits to their users (Serna et al., 2022).
- **Awareness of misuse and risks** in AI refers to the understanding and recognition of the potential for artificial intelligence systems to be used for harmful or unethical purposes, and the steps taken to prevent such occurrences (NIST, 2022). This can include considerations in the design, development, and deployment of AI systems to minimize the risk of misuse, as well as education and training for those who work with AI to ensure ethical and responsible use (IEEE, 2019).
- **Competence** in AI refers to the knowledge and skills required for safe and effective operations (IEEE, 2019). This knowledge should be specified by creators and operators should adhere to it.

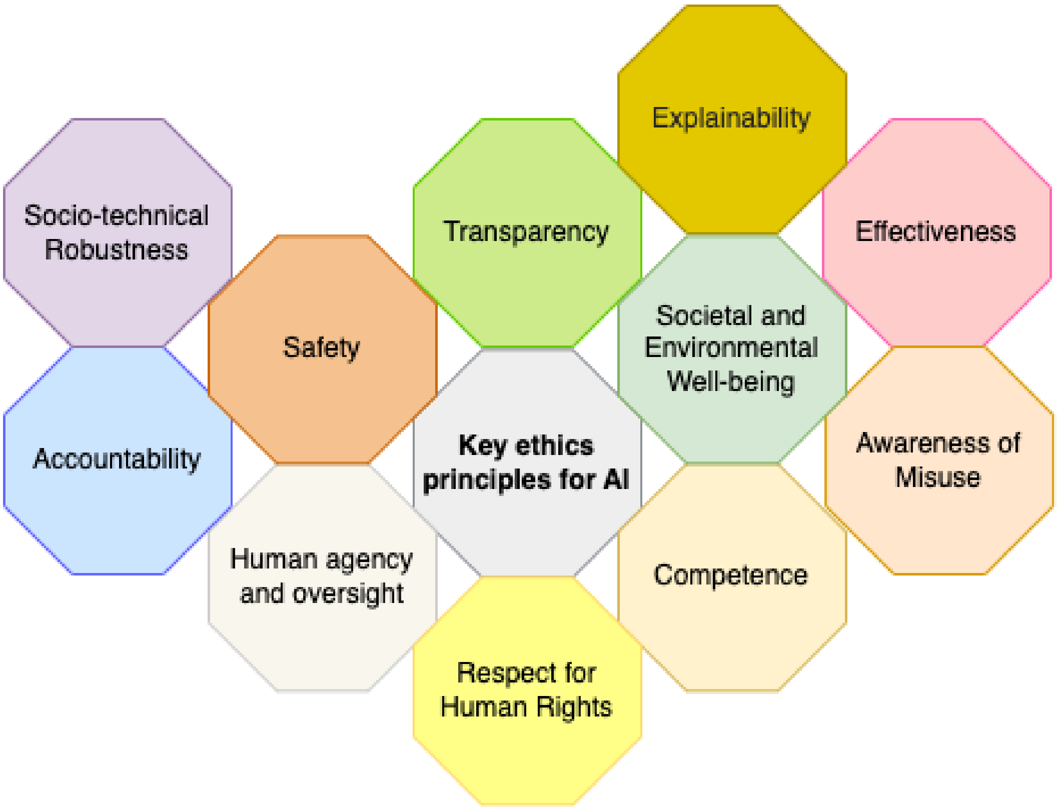

Explainability
Socio-technical Robustness
Transparency
Effectiveness
Safety
Societal and Environmental Well-being
Accountability
Key ethics principles for AI
Awareness of Misuse
Human agency and oversight
Competence
Respect for Human Rights

Figure 2: Key ethics principles for AI

3.3.3 Ethics in AI and Law Enforcement

The emphasis on utilizing AI in law enforcement is a relatively recent development, with guidelines and documents addressing the topic being published within the past 5 years. This field is still in the early stages of development and lacks significant detail. Below we summarize the most important documents.

3.3.3.1 The European Ethical Charter on the Use of AI in Judicial Systems

The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment was adopted by the European Commission for the Efficiency of Justice of CoE in 2018 (Council of Europe, 2018). This document sets out ethical principles for the use of AI in the judicial system. It aims to ensure that AI is used in a responsible and ethical manner, and that the rights and interests of individuals are protected. Some of the key principles set out in the charter include:

- **Respect of Fundamental Rights**: AI must be used in a way that is consistent with the rule of law and human rights, and that protects the rights and interests of individuals.
- **Quality and Security**: when processing judicial decisions and data, AI systems should use certified sources and data with models conceived in a multidisciplinary manner, in a secure technological environment.

- **Non-discrimination**: AI must be fair and non-discriminatory, and must not make decisions based on factors such as race, gender, religion, or other protected characteristics.
- **Transparency, impartiality and Fairness:** data processing methods should b accessible, and understandable and should authorise external audits
- **User under control**: it is vital to ensure that users are informed actors and in control of their choices.

The charter is intended to serve as a guide for policymakers, judiciary, and stakeholders involved in the development and use of AI in the judicial system, and to promote the responsible and ethical use of AI in this context.

3.3.3.2 Resolution 2342: Justice by Algorithm

The Resolution "Justice by Algorithm: the role of artificial intelligence in policing and criminal justice systems" by the Council of Europe stresses the need for accountability, transparency, and ethical considerations in the use of AI in the criminal justice system (Council of Europe, 2020). It also highlights the potential risks and challenges posed by AI, such as bias, discrimination, and the erosion of human rights, and calls for the development of appropriate legal frameworks and oversight mechanisms to ensure that AI is used in a way that is consistent with human rights standards. Additionally, the rresolution emphasizes the need for the responsible use of AI in policing and criminal justice, including ensuring that the use of AI is subject to human oversight, that it is transparent, and that individuals have the right to challenge decisions made using AI. The resolution specifically mentions as key principles:

- transparency, including accessibility and explicability.
- justice and fairness, including non-discrimination.
- human responsibility for decisions, including liability and the availability of remedies;
- safety and security
- privacy and data protection

It also refers to some specific use cases such as facial recognition, predictive policing, the identification of potential victims of crime, risk assessment in decision-making on remand, sentencing and parole, and identification of ‘cold cases’ that could now be solved using modern forensic technology.

3.4 Gaps

By analysing the content of the taxonomy outlined in the Appendix B we identified both general and specific gaps.

3.4.1.1 General gaps

- **Specificity:** The number of existing LEA-Guidelines is large, covering a large historical period from the mid-20th century until today. The number of AI-Guidelines is also large, covering a much shorter historical period from approximately 2010 until today. In contrast, the number of specific LEA/AI-Guidelines is small, confirming the importance of the current effort.

- **Locality:** Of the few LEA/AI-Guidelines in the taxonomy, most focus on US and UK localities. There are some international and European guidelines, but almost no general European sources, again confirming the value of the current effort. For the development of novel LEA/AI-Guidelines for the European region, it should be noted that although US and UK sources may be used for inspiration, there are socio-cultural differences to consider.
- **Algorithm vs. Application:** There is much focus on high-level "algorithmic ethics", and few sources consider the sector specific use of algorithms in law enforcement.
- **Principles vs. Operationalisation:** Significant efforts are underway to clarify ethical principles and to operationalize them through AI ethics tools and self-assessment checklists. However, much of this work has primarily focused on general ethical principles and their application in a general context. It is important to recognize that the law enforcement environment has its own unique characteristics and requirements that require specific considerations and adaptations. As a result, additional effort is needed to translate the general ethical principles into practical guidance for the law enforcement context, taking into account its particularities.
- **Limited functionalities:** The primary focus of the guidelines reviewed is on three specific areas of AI application: predictive policing, facial recognition, and chatbots. This focus is driven by a concern for fairness, bias, and discrimination (see values and functionalities in the taxonomy in Appendix B) in these applications. It is important to note that the guidelines have limited coverage of other functionalities listed in the functionalities taxonomy, as described in D2.1. Further research and analysis may be needed to address the ethical considerations of these other functionalities in a comprehensive manner.

3.4.1.2 Specific gaps

- **Safety** in Law Enforcement ethical guidelines is typically quantified through the absence of physical harm. However, in the context of LEAs and AI, it is crucial to also take into account non-physical harms such as violations or threats to individual autonomy or privacy. These non-physical harms are challenging to quantify and measure, and existing AI guidelines should be consulted to address this issue within the context of LEAs and AI.
- **Intervention.** With respect to LEAs ethical guidelines, interventions are traditionally measured in terms of physical interventions such as physical restraint or the seizure of physical property. AI technology, however, affords other kinds of interventions such as data-driven hyper nudges to influence an individual's or a group's behavior, or active crime-prevention measures stemming from the recommendations of data-driven predictive policing. A gap in current thinking about intervention in the LEA/AI context derives from a failure to consider the unique ways in which AI technology can be used to influence citizen's behavior via nonphysical interventions e.g. in the online domain.
- **Impartiality and Fairness.** Impartiality or bias is traditionally conceived as a feature of individual persons. Accordingly, existing LEA-Guidelines emphasize the needed of impartiality for law enforcement officers. However, the introduction of AI technology in LEA raises the possibility that the tools used in the LEA are biased. Unlike guns or vehicles, AI technology might itself be biased: the model might be based on biased datasets or built and used under

biased objective functions. Accordingly, LEA/AI-Guidelines should govern not only the use of AI technology by law enforcement agencies, but also the development and procurement of the technology itself.

- **Accountability.** LEAs ethics guidelines hold LEA personnel accountable to the state, citizens and their representatives for their actions and omissions. The use of opaque AI systems in decision-making may pose difficulties in determining the accountability of officers for their actions and omissions. Determining who is responsible for decisions and actions made by AI systems or based on AI systems raises questions about accountability, particularly when the process is not transparent and the decision making process is difficult to understand. Actions and omissions based on opaque AI systems might be particularly problematic to judge officers accountability.

4 Toward LEA/AI Guidelines

This section presents the outcome of the fourth research phase, which involved the creation of the popAI guidelines. These guidelines are a result of an extensive analysis and synthesis of the works included in taxonomy. The creation of the popAI guidelines represents an important step in advancing our understanding of ethical considerations in AI and the development of practical guidance for AI applications.

The PopAI LEA/AI guidelines are developed by combining the existing LEA guidelines and AI guidelines and are centered around two verticals: education and collaboration. The starting point is the current LEA guidelines (see Section 3), which are then modified or expanded to address identified gaps and include AI principles (see Section 3). The new LEA/AI Guidelines are accompanied by illustrative examples from the PopAI functionalities taxonomy, the controversies report and are paired with key recommendations that demonstrate their level of implementation.

4.1 (LEA/AI-a) Ensure and promote public safety, in both the physical and digital domains.

AI technology can be used by law enforcement agencies to ensure and promote public safety, as well as to reduce crime. This technology might be quicker, more precise, and more cost-effective than non-digital technologies. For example, face recognition techniques can be used in criminal investigations to quickly and accurately identify suspect individuals, as well as to efficiently and discreetly track their movements in public spaces. Similarly, data-driven predictive policing systems can identify criminal hotspots with higher spatiotemporal precision than traditional methods, allowing law enforcement agencies to deploy preventative measures more effectively and surgically. Also, social network analysis can be used to identify and track the online activity of criminal organizations, supplementing the traditional surveillance measures that are better suited for offline activities.

However, the use of AI technology also bears risks that can have a negative impact on public safety. For example, predictive policing systems are notoriously prone to bias (LEA/AI-c), possibly resulting in ineffective law enforcement, a loss of public trust, and excessive as well as unwarranted intervention (LEA/AI-b). Moreover, face-recognition and social network analysis used for surveillance can result in the unnecessary and excessive collection of data, leading to a loss of privacy. Although such negative effects may be difficult to detect and quantify, they bear the potential for significant harm to societal stability and individual rights.

For this reason, although AI technology can be an effective tool for ensuring public safety and preventing crime, its use must be measured and balanced against the possible risks to preserve the safety of all citizens.

Measuring and balancing AI risks in policing can be achieved through the following steps:

- **Identify potential risks**: Assess the potential negative impacts of AI in policing, such as privacy violations, racial biases, and error-prone predictions.
- **Evaluate algorithms**: Regularly evaluate the algorithms used in policing through qualitative and quantitative metrics to ensure they are fair, transparent, and accurate.
- **Engage stakeholders**: Involve various stakeholders, including law enforcement agencies, community groups, and social, cultural and technical experts, in the development and implementation of AI in policing.
- **Educate:** organise socio-technical training sessions with law enforcement officers on computer and coding skills and AI benefits and risks.
- **Establish clear policies**: Develop clear policies and procedures for using AI in policing, including ethics, data protection and privacy, transparency, and accountability.
- **Monitor outcomes**: Continuously monitor and assess the outcomes of AI in policing to identify any adverse impacts and to make necessary changes.
- **Foster transparency**: Foster transparency in AI decision-making processes to increase trust and accountability.
- **Encourage accountability**: Hold law enforcement agencies accountable for the impacts of AI on individual rights and liberties and enforce accountability measures.

4.2 (LEA/AI-b) Restrain AI intervention and preserve autonomy

Like many non-digital technologies, AI technologies can be used to influence and intervene on citizens' behaviours, thereby possibly preempting or combatting criminal activity. For example, image- and video-classification technology can be used to identify copyrighted or illegal material and prevent its dissemination over the internet via upload-filters. Moreover, data-driven hypernudges can be used to prevent criminal activity by influencing citizen behaviour in public spaces. These and other AI-driven applications can help law enforcement agencies ensure public safety and facilitate crime prevention (LEA/AI-a).

At the same time, these uses of AI technology may threaten citizens' autonomy, that is, citizens' ability to make informed, uncoerced decisions in their everyday activities. For example, upload-filtering limits citizens' ability to create and share digital content as they please, and hypernudging influences citizen behavior in ways that they might not desire or understand. Law enforcement agencies have the duty of minimizing such threats to citizen autonomy.

In order to ensure that autonomy is preserved, AI-driven interventions must be used only when the relevant law-enforcement goals cannot be achieved by other means, e.g. through the use of traditional technologies. Moreover, when the use of AI technology is deemed necessary, its use must be indicated clearly and transparently (LEA/AI-d).

Restraining AI intervention in policing while preserving autonomy can be achieved through the following steps:

- **Human oversight**: Implement human oversight of AI systems in policing to ensure they are used in a responsible and ethical manner.
- **Limiting AI's role:** Limit the role of AI in policing to specific tasks or circumstances.
- **Transparency**: Ensure transparency in AI decision-making processes and make algorithms open to scrutiny.
- **Independent audits**: Conduct regular independent audits of AI systems in policing to identify and address any biases or inaccuracies.
- **Educate:** organise socio-technical training sessions with law enforcement officers on computer and coding skills and AI benefits and risks.
- **Regular evaluations**: Regularly evaluate the outcomes of AI systems in policing to ensure they are aligned with human values and ethical principles.
- **Ethical guidelines**: Adopt ethical guidelines for the use of AI in policing to ensure that individual rights and liberties are protected.
- **Collaborative governance**: Foster collaboration between law enforcement agencies, communities, and socio- technical experts to ensure that AI is used in a manner that serves the public interest.

4.3 (LEA/AI-c) Avoid AI bias, and promote Impartial and fair treatment of individuals

All citizens are equal before the law and law enforcement agencies have the duty of treating individuals impartially and fairly. This duty extends to the technologies used in law enforcement, including AI technology.

AI technology has been known to be biased. For example, predictive policing systems have been known to disproportionally affect citizens of underrepresented and historically disenfranchised groups, and face-recognition systems have been known to have lower accuracy for certain populations than for others. As a consequence, the use of AI technology in law enforcement can have different effects on different citizens, and lead to unfair outcomes.

In order to achieve the goal of impartial and fair treatment for all, law enforcement should ensure that the AI technologies and their use are unbiased. This mandate should influence the development of relevant technologies and should be considered by law enforcement agencies during procurement and use.

Guaranteeing fair treatment of individuals when using AI in policing can be achieved through the following steps:

- **Unbias systems**: ensure that AI algorithms used in policing are tested and trained to unbias the systems, particularly in regard to race, gender, age, socio-economic status and other protected characteristics.

- **Unbias use**: ensure that the use of AI algorithms does not target specific populations in regard to race, gender, socio-economic status or other protected characteristics.
- **Educate**: LEAs should mandate the documented use of state-of-the-art bias-identification and bias-mitigation techniques.
- **Explainable AI**: Use "Explainable AI" (XAI)[4] technologies to make AI decision-making processes transparent and understandable to humans.
- **Human oversight**: Implement human oversight of AI systems in policing to ensure they are used in a responsible and ethical manner.
- **Regular audits**: Conduct regular independent audits of AI algorithms used in policing to identify and address any biases or inaccuracies.
- **Open data**: Encourage open data practices in policing to increase transparency and accountability in AI decision-making.
- **Stakeholder engagement**: Involve diverse communities and other stakeholders in the development and deployment of AI systems in policing to ensure that their perspectives and concerns are taken into account.

4.4 (LEA/AI-d) Transparency and training to achieve Integrity and respect for the rule of law when using AI

In order to respect the rule of law and restrict LEAs arbitrary power, AI systems that are used or planned to be used by law enforcement officers need to be understood by their users. Recent Data Protection Authorities' decisions (see D2.2 for a review) show that Law Enforcement officers' lack of training in AI led to breaches. Data laws have been violated by Swedish LEAs using a facial recognition software without conducting a data protection impact assessment (DPIA) and failing to set appropriate measures to ensure that the data processing was in accordance with the law. The Swedish data watchdog ordered the police further training and education to avoid breaches (European Data Protection Board, 2021).

Guaranteeing integrity and respect for the rule of law when using AI in policing can be achieved through the following steps:

- **Adequate training**: Provide adequate training to law enforcement agencies on the use of AI to ensure they understand the technology, its limitations, legal requirements and ethical considerations.
- **Clear policies**: Develop clear policies and procedures for the use of AI in policing, including guidelines for data protection and privacy, non-discrimination, transparency, and accountability.
- **Human oversight**: Implement human oversight of AI systems in policing to ensure they are used in a manner that respects the rule of law.

[4] See for example: https://xai-project.eu

- **Regular evaluations:** Regularly evaluate the outcomes of AI systems in policing to ensure they align with human values and ethical principles.
- **Transparency**: Ensure transparency in AI decision-making and make algorithms open to scrutiny.
- **Independent audits**: Conduct regular independent audits of AI systems used in policing to identify and address any biases or inaccuracies.
- **Collaborative governance:** Foster collaboration between law enforcement agencies, communities, and socio-technical experts to ensure that AI is used in a manner that serves the public interest and respects the rule of law.

4.5 (LEA/AI-e) Protect and promote individual rights when using AI

LEAs' use of AI must respect human rights. Recent Data Protection Authority decisions show that this is a real challenge from multiple standpoints. The recent European Data Protection Supervisor (EDPS)-EURPOL case is a good example which raises important concerns. In January 2022, the EDPS ordered Europol to delete data of individuals with no link to criminal activity to protect citizens' fundamental right to privacy (European Data Protection Supervisor, 2022). As a response, EUROPOL amended its regulation, rendering ineffective the EDPS decision and expanding Europol capacity to exchange personal data with third parties, the use of AI, and the processing of large datasets Supervisor (European Data Protection Supervisor, 2022). In this structure of power, the balance between security and the protection of the rights of those who are subjected to the use of AI must be taken seriously.

For example, surveillance, communication, and recognition functions powered by AI tools might affect individuals behaviorally and psychologically and can push them to conform to certain norms, putting at stake human dignity and the right to the Integrity of the person. AI powered systems that automatically track individuals and their communication might jeopardise their rights to freedom of expression and assembly. Because of surveillance, people might be discouraged to participate to social protests and demonstrations and might diminish their willingness to express their opinion. AI systems are often seen as neutral and objective but they replicate human biases infringing people right to non-discrimination. The risk of discrimination can arise from biased data, from biased design of the algorithm or optimisation, or from its contextual use. For example, facial recognition has been shown to have higher error rates for women and women of colour, as well as for data analytics (Buolamwini & Gebru, 2018). Another example is the data used for predictive policing which has been demonstrated to reflect patrolling priorities in disadvantaged areas rather than criminal activities, risking to lead to a self-fulfilling prophecy. Furthermore, AI systems use, track, and recognize data to inform law enforcement actions and judiciary decisions. The opaqueness that characterises the AI systems makes it hard to understand the reasoning behind an output, to challenge the decision and have an effective remedy. This has a direct effect on people's right to have a fair trial, their presumption of innocence and their right to have an effective remedy. The use of AI by LEAs creates also new risks and challenges for Social and Economic rights of the persons working in law enforcement and judiciary environment. For example, the use of AI implies new training for LEAs and judiciary authorities, and new assessment for risks of their use at the workplace.

Ensuring that AI in policing protects and promotes individual rights can be achieved through the following steps:

- **Data protection and privacy**: Conduct a Data protection Impact assessment to evaluate whether the technology under development or its use adhere to Data Protection Laws.
- **Human rights:** Conduct a Human Rights Impact assessment to evaluate the effect of the AI technology or its use on human rights.
- **Human oversight**: Implement human oversight of AI systems in policing to ensure they are used in a responsible and ethical manner.
- **Transparency**: Ensure transparency in AI decision-making processes and make algorithms open to scrutiny.
- **Fairness and non-discrimination**: Ensure that AI algorithms used in policing are tested and trained to reduce biases and promote fairness and non-discrimination.
- **Regular evaluations**: Regularly evaluate the outcomes of AI systems in policing to ensure they align with human values and ethical principles.
- **Stakeholder engagement**: Involve diverse communities and other stakeholders in the development and deployment of AI systems in policing to ensure that their perspectives and concerns are taken into account.
- **Independent audits**: Conduct regular independent audits of AI systems used in policing to identify and address any biases or inaccuracies.
- **Ethical guidelines**: Adopt ethical guidelines for the use of AI in policing to ensure that individual rights and liberties are protected.
- **Legal compliance**: Ensure that AI systems used in policing comply with all relevant laws, regulations, and ethical principles.

4.1 (LEA/AI-f) Community AI policing: Serve while building relationships

Previous cases of technology adoption by law enforcement agencies have resulted in distrust and rejection among the community due to a lack of clear explanation regarding the technology's intended use and scope. In 2010, the installation of CCTVs in a Muslim area in Birmingham, UK caused controversy (Lewis, 2010). People initially believed the cameras were for tracking drivers and protesters, but it was later revealed that the cameras were installed to monitor suspected extremists among the city's Muslim population. The cameras were installed without consultation and included 150 ANPR cameras, 40 of which were classified as "covert". The initiative was sponsored by the Terrorism and Allied Matters fund, and the cameras were positioned to track anyone entering or leaving the neighbourhood. The community felt misled and targeted as suspected terrorists. Briefing documents given to councillors played down the importance of counterterrorism, and only mentioned it as a secondary benefit of the cameras, stating that they would provide support and reassurance to communities considered vulnerable to violent extremism.

This case highlights the importance of effectively communicating and engaging with the community prior to introducing new technologies. Community engagement is essential for the effective deployment of AI in law enforcement activities. It serves to foster mutual understanding and trust regarding the technology use, enabling the development of a constructive dialogue and promoting ethical applications of AI. Through community engagement, misunderstandings can be prevented, and the implementation of AI can be aligned with the community's values and priorities.

Applying community policing to AI in law enforcement can involve the following steps:

- **Stakeholder engagement:** Engage with community members and other stakeholders to understand their perspectives on the use of AI in policing and involve them in the decision-making process.
- **Transparency**: Ensure transparency in AI decision-making processes and make algorithms open to scrutiny.
- **Human oversight**: Implement human oversight of AI systems in policing to ensure they are used in a responsible and ethical manner.
- **Collaborative problem-solving**: Encourage collaboration between police officers and community members to identify and address public safety problems using a combination of traditional policing methods and AI technologies.
- **Partnership building**: Build partnerships between the police and community organizations to enhance public safety and improve community well-being.
- **Community policing training:** Provide training for police officers on community policing principles and practices to ensure they are equipped to engage with community members in a positive and productive manner.
- **Regular evaluations and report**: Regularly evaluate and report to the public the outcomes of AI systems in policing to ensure they align with community values and are effectively addressing public safety problems.
- **Community feedback**: Encourage community members to provide feedback on the use of AI in policing and incorporate their suggestions and concerns into ongoing decision-making. It is important to engage individuals with diverse backgrounds and sociocultural values in an inclusive manner in order to incorporate a wider variety of views and concerns.
- **Ethical guidelines**: Adopt ethical guidelines for the use of AI in policing that prioritize community well-being and individual rights and liberties.

5 Conclusion

The integration of AI in policing holds the promise of improved efficiency and effectiveness through identifying patterns, automating tasks, and streamlining processes. However, it is vital to weigh the potential benefits against the risks that AI can pose to human rights and democracy, including the perpetuation of bias and potential misuse of the technology. Further, AI use may erode public trust in law enforcement and exacerbate social issues. Hence, proper procedures must be in place to address these concerns and ensure responsible implementation of AI in policing.

The objective of this report was to provide law enforcement agencies with a comprehensive understanding of the ethical considerations surrounding the implementation of artificial intelligence (AI) in policing. The report aimed to:

- Define the concept of ethics and its relevance to AI applications in law enforcement
- Highlight the ethical risks associated with various AI applications in policing
- Help LEAs navigate the complex field of AI ethics for LEAs by creating a taxonomy
- Propose a set of guidelines for LEAs to adhere to in their ethical use of AI.

To achieve this, section 2 defined ethics and AI ethics, and described the objectives of ethical mechanisms. It also analyzed the key ethical considerations involved in the use of AI for crime prevention and investigation, migration management, administration of justice, cyber operations, and LEA training. Section 3 presented the methodology employed to develop the PopAI taxonomy and guidelines. The key ethics frameworks and principles relevant to law enforcement agencies (LEAs) and AI were discussed in detail as well as their gaps. Section 4 outlined the PopAI guidelines for LEAs, which have been created through a combination of existing LEA guidelines and AI principles. The process involved modifying and expanding upon the current LEA guidelines to address identified gaps and incorporate principles specific to AI. The guidelines are accompanied by illustrative examples drawn from the PopAI functionalities taxonomy and controversies report, and by key recommendations to demonstrate the level of implementation.

In conclusion, the integration of ethical considerations regarding AI usage within law enforcement helps mitigate potential negative consequences, minimize risks, and foster public confidence. It is crucial to create a culture of AI ethics, implement training programs, engage the public, listen to civil societies and actively participate in discussions and developments within the AI socio-technical communities to enhance the overall trustworthiness of AI applications in law enforcement.

6 References

AI HLEG. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

AI HLEG. (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Al-Jarani, Y. (2019). All Fun and (Mind) Games? Protecting Consumers from the Manipulative Harms of Interactive Virtual Reality. *Ill. J.L. Tech. & Pol'y*. https://heinonline.org/HOL/LandingPage?handle=hein.journals/jltp2019&div=16&id=&page=

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Anthes, C., Garcia-Hernandez, R. J., Wiedemann, M., & Kranzlmuller, D. (2016). State of the art of virtual reality technology. *2016 IEEE Aerospace Conference*, 1–19. https://doi.org/10.1109/AERO.2016.7500674

Bechara, A., Damasio, H., & Damasio, A. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex*, *10*(3), 295–307. https://doi.org/10.1093/cercor/10.3.295

Beduschi, A. (2022). Migration and Artificial Intelligence. In *Migration and artificial intelligence.* Routledge.

Bircan, T., & Korkmaz, E. E. (2021). Big data for whose sake? Governing migration through artificial intelligence. *Humanities and Social Sciences Communications*, *8*(1), 241. https://doi.org/10.1057/s41599-021-00910-x

Brenkert, G. G. (2009). Google, Human Rights, and Moral Compromise. *Journal of Business Ethics*, *85*(4), 453–478. https://doi.org/10.1007/s10551-008-9783-3

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 1:15. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Cahn, S. M., & Markie, P. (2019). *Ethics: History, Theory, and Contemporary Issues*. Oxford University Press.

Coeckelbergh, M. (2020). *AI Ethics*. The MIT Press Essential Knowledge Series.

Council of Europe. (2001a). *Recommendation Rec(2001)10 of the Committee of Ministers to member states on the European Code of Police Ethics*. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016804c1a0c

Council of Europe. (2001b). *The European Code of Police Ethics*. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016804c1a0c

Council of Europe. (2018). *CEPEJ European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment#:~:text=The%20CEPEJ's%20view%20as%20set,on%20Human%20Rights%20(ECHR)%20and

Council of Europe. (2020). *Justice by algorithm—The role of artificial intelligence in policing and criminal justice systems*. https://pace.coe.int/en/files/28805/html

Driver, J. (2006). *Ethics: The Fundamentals*. Blackwell Publishing.

EDRi. (2022). *News from Ireland question effectiveness and lawfulness of online scanning for tackling child sexual abuse: Lessons for the EU*. https://edri.org/our-work/breaking-irish-story-shows-that-eus-csam-proposal-can-never-work/

European Data Protection Board. (2021). *Swedish DPA: Police unlawfully used facial recognition app*. https://edpb.europa.eu/news/national-news/2021/swedish-dpa-police-unlawfully-used-facial-recognition-app_en

European Data Protection Supervisor. (2022). *EDPS takes legal action as new Europol Regulation puts rule of law and EDPS independence under threat*. https://edps.europa.eu/press-publications/press-news/press-releases/2022/edps-takes-legal-action-new-europol-regulation-puts-rule-law-and-edps-independence-under-threat_en

EUROPOL. (2019). *The Code of Conduct of Europol*. https://www.europol.europa.eu/sites/default/files/documents/the_code_of_conduct_of_europol.pdf

Fair Trials. (2021). *Automating Injustice: The use of artificial intelligence & automated decision making systems in criminal justice in Europe*. Fair Trials. https://www.fairtrials.org/articles/publications/automating-injustice/

Fair Trials. (2022). *Regulating Artificial Intelligence for Use in Criminal Justice Systems in the EU*. https://www.fairtrials.org/app/uploads/2022/01/Regulating-Artificial-Intelligence-for-Use-in-Criminal-Justice-Systems-Fair-Trials.pdf

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Future of life Institute. (2017). *Asilomar AI Principles*. https://futureoflife.org/open-letter/ai-principles/

IACP. (2019). *Standards of Conduct*. https://www.theiacp.org/sites/default/files/2020-06/Standards%20of%20Conduct%20June%202020.pdf

IEEE. (2019). *Ethically aligned design*. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

King, O. C., & Mertens, M. (2023). Self-fulfilling Prophecy in Practical and Automated Prediction. *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-022-10359-9

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

Lewis, P. (2010). Surveillance cameras spring up in Muslim areas-the targets? Terrorists. *The Guardian*. https://www.theguardian.com/uk/2010/jun/04/birmingham-surveillance-cameras-muslim-community

Manners, I. (2008). *The Normative Ethics of the European Union*. *84*(1), 45–60.

Molnar, P. (2020). *Technological Testing Grounds. Migrations Management Experiments and Reflections from the Ground Up.* EDRi, Regugee Law Lab. https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf

Molnar, P. (2021). Robots and refugees: The human rights impacts of artificial intelligence and automated decision-making in migration. In *Research Handbook on International Migration and Digital Technology* (pp. 134–151). Edward Elgar Publishing. https://www.elgaronline.com/display/edcoll/9781839100604/9781839100604.00019.xml

NIST. (2022). *AI Risk Management Framework*. https://www.nist.gov/itl/ai-risk-management-framework

Northern Ireland Policing Board. (2008). *Police Service of Northern Ireland Code of Ethics 2008*. https://www.nipolicingboard.org.uk/files/nipolicingboard/publications/code-of-ethics.pdf

OECD. (2019). *OECD AI Principles*. https://oecd.ai/en/ai-principles

Park, C. S. Y., Haejoong, K. I. M., & Sangmin, L. E. E. (2021). Do Less Teaching, Do More Coaching: Toward Critical Thinking for Ethical Application of Artificial Intelligence. *Journal of Learning and Teaching in Digital Age*, *6*(2), 97–100.

Rességuier, A., & Rodrigues, R. (2020). *AI ethics should not remain toothless!* A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), 205395172094254. https://doi.org/10.1177/2053951720942541

Rubio, O., Estella, A., Cabre, L., Saralegui-Reta, I., Martin, M. C., Zapata, L., Esquerda, M., Ferrer, R., Castellanos, A., Trenado, J., & Amblas, J. (2020). Ethical recommendations for a difficult decision-making in intensive care units due to the exceptional situation of crisis by the COVID-19 pandemic: A rapid review & consensus of experts. *Medicina Intensiva (English Edition)*, *44*(7), 439–445. https://doi.org/10.1016/j.medine.2020.06.002

Schuller, D., & Schuller, B. W. (2018). The Age of Artificial Emotional Intelligence. *Computer*, *51*(9), 38–46.

Serna, I., Morales, A., Fierrez, J., & Obradovich, N. (2022). Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning. *Artificial Intelligence*, *305*, 103682. https://doi.org/10.1016/j.artint.2022.103682

Sinclair, A. (1993). Approaches to organisational culture and ethics. *Journal of Business Ethics*, *12*(1), 63–73. https://doi.org/10.1007/BF01845788

Singer, P. (2011). *Practical Ethics* (Third Edition). Cambridge University Press.

Singleton, A. (2016). *Migration and Asylum Data for Policy-Making in the European Union. The Problem with Numbers.* CEPS. https://www.ceps.eu/ceps-publications/migration-and-asylum-data-policy-making-european-union-problem-numbers/

Taddeo, M. (2013). Cyber Security and Individual Rights, Striking the Right Balance. *Philosophy & Technology*, *26*(4), 353–356. https://doi.org/10.1007/s13347-013-0140-9

Taddeo, M. (2019). Three Ethical Challenges of Applications of Artifcial Intelligence in Cybersecurity. *Minds and Machines*, *29*, 187–191.

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 272–283. https://doi.org/10.1145/3351095.3372834

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

United Nations. (1979). *Code of Conduct for Law Enforcement Officials*. https://www.ohchr.org/en/instruments-mechanisms/instruments/code-conduct-law-enforcement-officials

United Nations. (1990). *Basic Principles on the Use of Force and Firearms by Law Enforcement Officials*. https://www.ohchr.org/en/instruments-mechanisms/instruments/basic-principles-use-force-and-firearms-law-enforcement#:~:text=Law%20enforcement%20officials%2C%20in%20carrying,of%20achieving%20the%20intended%20result.

Van den Hoven, J., Lokhorst, G.-J., & Van de Poel, I. (2012). Engineering and the Problem of Moral Overload. *Science and Engineering Ethics*, *18*(1), 143–155. https://doi.org/10.1007/s11948-011-9277-z

Weinstock, D. (2013). On the possibility of principled moral compromise. *Critical Review of International Social and Political Philosophy*, *16*(4), 537–556. https://doi.org/10.1080/13698230.2013.810392

Zardiashvvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement. *Delphi*, *4*, 1–7.

7 Appendix A

1.AI

- What is AI?
- What are the current capabilities for policing? In which areas? Where?

2.Ethics and AI

- What in detail does 'ethics' mean?
- Who defines what ethics is?
- Is there any fixed definition of ethics and of how to behave ethically (in policing)?
- How are AI and ethics related to each other? Are there any examples?
- When and how would the use of certain software (including AI) in policing be ethically ok or not ok?
- Doesn't our legislation completely reflect ethical standards?
- To what extent should ethical issues be considered beyond the legal framework?
- What are the most important principles in AI ethics?
- How can an AI system be used ethically and not depend on the ethics of the end user?
- Ethics is about doing the right thing, but with all upcoming acts and assessments, the focus seems to be on doing things right. How can we make sure we're doing the right thing instead of focusing on doing things right?
- Another question is one about the ethical requirements for trustworthy AI within the Law Enforcement. The EC has seven (ethical) requirements for trustworthy AI. How do these requirements apply within the LEA and are there other ethical requirements specifically for the LEA?
- Why it is important for technical partners to think about ethics if they are developing a tool which is used by LEAs, specially that the usage of this tool will require anyhow a legal evaluation of the data protection legal acts/court act, before they are allowed to use it.
- Is it unethical to not use AI if it is possible to use it?

3.Bias & specific principles

- How could AI bias be avoided?
- Is the system designed for traceability, accountability and auditability?
- Does the system allow for human supervision and monitoring?

4.AI & police organisational needs

- How well does AI meet the needs of the police?
- Should the police be given specific training on A.I.?
- Should there be plans on prospective in Police organizations as to A I.?
- How could LEAs adapt to the constant technological evolution in A.I.?

- What repercussions would have A.I. in police work in terms of effectiveness and efficiency? what is the impact of Artificial intelligence on police organizations short and long term?
- How could it be ensured that the use of an AI system fulfils the purposes of the law enforcement agency while being within an ethical framework?

5.AI police and society

- Who should mark the technological development of A I.? society? (including LEAs) or/and technology companies?
- What is the impact of A.I. on citizens and their opinion on the use of A.I. by the police?
- From the citizen's perspective: can artificial intelligence be seen as a control tool or a friendly technology?
- To what extent can AI help or hinder?

8 Appendix B

To access the PopAI ethics taxonomy click here or go to

https://docs.google.com/spreadsheets/d/17FwmU5dmMcc5indT8VplvCzBAZvxd1XdS6w64PRIBr8/edit?usp=sharing