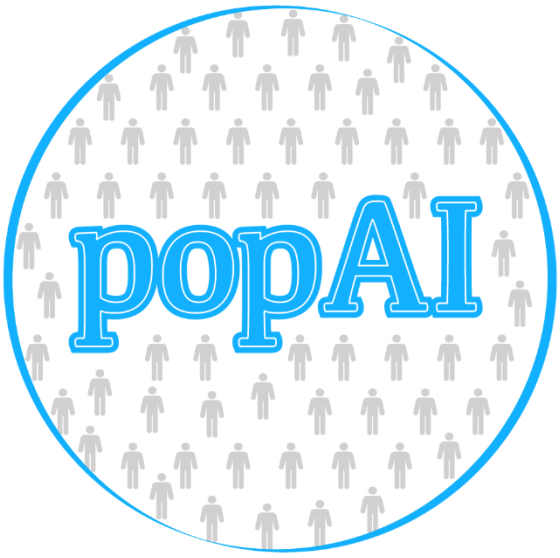

A European Positive Sum Approach towards AI tools in support of Law Enforcement and safeguarding privacy and fundamental rights

D5.8: popAI roadmaps

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	30/09/2023		
Nature	R = Document, report	**Dissemination Level**	PU = Public
Authors	Paola Fratantoni, Domenico Frascà (Z&P)		
Contributors	Pinelopi Troullinou, Francesca Trevisan (TRI) Elena Galifianaki, Dimitris Kyriazanos (NCSRD) Dimitra Papadaki (KEMEA) Anthoula Bania (HP)		
Reviewers	Andreas Ikonomopoulos, Vangelis Karkaletsis (NCSRD)		

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

Executive Summary

This report includes the findings of task 5.5 *popAI roadmaps*, carried out in the second half of the project. This task is divided into two subtasks, both addressed in this deliverable.

Subtask 5.5 (a) *popAI compliance and certification roadmap* aims to define a short-term roadmap, focusing on the achievement of a "European common approach" for compliance and certification of AI-based technologies used in the frame of Law Enforcement activities. This short-term roadmap is addressed in the first part of the deliverable (Part A).

This subtask was carried out conducting an intensive analysis of the current national, European and international documents that set standards, guidelines and a regulatory framework for AI, as well as research and policy papers. This work allowed to gain a more comprehensive understanding of the current landscape, while paving the way for a gap analysis, aiming at identifying areas where a stronger normative framework is needed. This gap analysis helped to formulate the final roadmap for compliance and certification of AI technologies.

Subtask 5.5 (b) *popAI Roadmap for 2040* provided a long-term roadmap, highlighting potential futures scenarios and risks, trying to understand the role that AI-based technologies might play in these futures and identifying strategies to get there. This roadmap emerged from a combination of different activities. Initially, a literature review was conducted to define the current European security landscape. This activity helped to understand the main security challenges and priorities set by the European Union. In addition, the worked carried out in WP3 (the foresight scenarios) has been embraced and further developed, by means of an interactive workshop with the stakeholder advisory board members. The outcomes from WP4 have also been relevant for the definition of the roadmap, which was drafted and discussed by the consortium members and presented in popAI final event, at the presence of popAI partners (including LEAs), sibling and related projects, popAI Project Officer and other representatives from the European Commission.

Table of Contents

Part A – Subtask (a) – popAI compliance and certification roadmap 6
1 Introduction 6
1.1 Background 6
1.1.1 The need for AI governance 6
1.1.2 The ALTAI principles 7
1.1.3 The EU AI Act Proposal 8
1.2 Definition and terminology 9
1.3 Structure of the deliverable and relation to other Work Packages 10
2 Methodology 12
3 Analysis of the State-of-the-Art of Standardisation and Certification standards 13
4 Definition of the roadmap 28
Part B – Subtask (b) – popAI Roadmap for 2040 33
5 Introduction 33
5.1 Background 33
5.1.1 The EU AI Strategy 33
5.2 Definitions and terminology 35
5.3 Structure of the deliverable and relation to other WPs 35
6 Methodology 37
7 The European security landscape 39
8 Inputs from WP3 and WP4 42
8.1 Input from WP3 foresight scenarios 42
8.2 Input from WP4 Recommendations for the ethical use of AI by LEAs 42
9 Definition of the roadmap 43
9.1 Workshop with Stakeholder Advisory Board 43
9.1.1 Purpose and structure of the workshop 43
9.1.2 Results of the workshop 44
9.2 The popAI 2040 roadmap 52
10 Conclusions 58
11 References 59
Annex I (Part A) 61
Annex II (Part A) 61

Annex III (Part B) 64
Annex III.1 Scenario 1 “Past will always define future” 64
Annex III.2 Scenario 2 “AI investigator. Case closed” 65
Annex III.3 Scenario 3 “Don’t shoot the artist” 65
Annex III.4 Scenario 4 “Crossing the invisible borders” 67
Annex III.5Scenario 5 “Guilty till proven innocent” 68
Annex IV (Part B) 69
Annex IV.1 69
Recommendations for LEAs 69
Annex IV.2 Recommendations for policymakers 69
Annex IV.3 Recommendations for technology developers 70

List of Figures

Figure 1 - Methodology for task 5.5 (a) 12
Figure 2 - popAI certification and compliance roadmap 29
Figure 3 - Milestones towards the EU AI Act 34
Figure 4 - Methodology for task 5.5 (b) 37
Figure 5 - Global risks landscape: an interconnections map (Source: [17]) 40
Figure 6 - EU Security Union Strategy: priorities (Source: [18]) 41
Figure 7 - Key output of the foresight scenarios workshop (Athens) 52
Figure 8 - popAI Roadmap to 2040 53

List of Tables

Table 1 - ALTAI Principles 7
Table 2 - popAI Standards Catalogue 14
Table 3 - popAI certification and compliance roadmap (steps) 30
Table 4 - Main sources used for the desk research 39
Table 5 - popAI Roadmap to (steps) 53

List of Terms & Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AI HLEG	High-Level Expert Group on Artificial Intelligence
ALTAI	
CEN	European Committee for Standardisation
CENELEC	European Electrotechnical Committee for Standardization
EC	European Commission
EN	European Standards
ESO	European Standardisation Organisation
ESO	European Standardisation Organisation
ETSI	European Telecommunications Standards Institute
EU	European Union
IEC	International Electrotechnical Commission
IRTF	Internet Research Task Force
ISO	International Organisation for Standardisation
ITU	International Telecommunication Union
LEA	Law Enforcement Agency
SDO	Standard-development Organisation
SSO	Standard-setting Organisation
W3C	World Wide Web Consortium
WP	Work Package

Part A – Subtask (a) – popAI compliance and certification roadmap

1 Introduction

Part A encapsulates the results of the research conducted under subtask (a), aiming to investigate the compliance and certification landscape of AI technologies. Sub-task (a) of T5.5, indeed, aiming to establish a *"'European common approach' for compliance and certification of AI technologies in support of Law Enforcement […]".* In order to achieve this objective, this subtask delineates a short-term *"compliance and certification roadmap"*, which illustrates specific actions for LEAs to be implemented as well as additional actions for policy officers. This short-term roadmap stems from an analysis of the existing legal framework to understand the certification landscape concerning AI.

It should be noted that a few challenges arose during the implementation of this research. Among these, the lack of a common consensus between what is AI and what is not AI; second, the different levels of regulation and systematisation of legal concepts between the private and public sectors and third, the lack of specific provisions dedicated to AI in the security domain. Last, the inconsistency in the legislative approaches among the EU MSs which makes it more difficult to achieve a common understanding of the concept.

Part A provides a preliminary overview of the ALTAI principles (Assessment List for Trustworthy Artificial Intelligence)[1] representing the backbone of the EU attempt to raise awareness about AI and the impact on the individual and the society as well as harmonise concepts, provide a framework to assess potential risks related to the use of such technologies in the different domains and actions to mitigate or minimise such risks.

This initial background will help the familiarisation with the concept and terminology; subsequently, an analysis of national, European and international frameworks and certifications has been conducted, thus allowing to gain a more comprehensive perspective of the current legal framework. This information will be organised in a catalogue to facilitate navigation into these frameworks.

The final section of part A includes the short-term roadmap, indicated actions and target audience.

1.1 Background

1.1.1 The need for AI governance

Artificial Intelligence (AI) is fundamentally transforming society, the economy and the way we live and work. The pervasive impact of digitalization extends across almost every domain of the society, encompassing fields such as medical care, road traffic, security, strategy, defence, and even the very nature of human communication. The rapidity at which these changes occur is unparalleled in history. As we endeavour to harness the potential of AI, it is equally imperative to exercise caution and uphold human rights and ethical principles.

As AI continues its expansion, it becomes paramount to strike a delicate balance between innovation and safeguarding the well-being of individuals and society. By adopting a responsible approach to AI development and deployment, it would be easier to ensure that its transformative power aligns with

[1] Please refer to section 1.1.2 for additional information on the ALTAI principles.

ethical standards and respects fundamental human rights. Guidelines need to be provided to facilitate the responsible and trustworthy use of AI. These guidelines should address ethical and legal issues to ensure AI is deployed in a manner that upholds societal values and protects individual rights, while providing countries a common framework to refer to.

Rules governing technology must be centred around human welfare to foster a sense of trust among individuals, assuring them that the implementation of technology adheres to safety standards and legal requirements, while respecting their fundamental rights. [1] [2] [3].

The certification and compliance roadmap emerges in this context, where the need for a common framework to certificate AI-based technologies and ensure that their application does not contravene ethical and legal principles is strongly needed. Such a common framework should facilitate the harmonisation of procedures across Europe when dealing with the use of AI technologies in security. The need for AI governance stems from the consideration that AI is constantly increasing its relevance in daily life and task, but still lacking a clear and common framework for its regulation. The roadmap presented in part A of this deliverable aims to complement the activities and initiatives promoted at the EU level in order to set and implement the EU AI Act, the first global regulatory framework for AI.

1.1.2 The ALTAI principles

In April 2019 the High-Level Expert Group on Artificial Intelligence (AI HLEG)[2]\ has published the **Ethics Guidelines** for Trustworthy Artificial Intelligence, 2019) in order to promote Trustworthy AI [2].

Trustworthy AI has three important components: *it should be **lawful** (complying with all applicable laws and regulations), it should be **ethical** (ensuring adherence to ethical principles and values and it should be **robust** both from a technical and social perspective.*

The Guidelines identify the ethical principles and how these must be respected in the development, deployment and use of AI systems (respect for human autonomy, prevention of harm, fairness and explicability), while providing guidance on how Trustworthy AI can be realised, by listing seven requirements that AI systems should meet (Table 1). Both technical and non-technical methods can be used for their implementation. In addition, the guidelines provide a concrete and non-exhaustive Trustworthy AI assessment list.

Table 1 - ALTAI Principles

Principle	Description
Human Agency and Oversight	AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights
Technical Robustness and Safety	AI systems need to be resilient and secure
Privacy and Data Governance	Full respect for privacy and data protection, safeguard data's quality and integrity, ensure legitimisation of data's access

[2] The AI HLEG is an independent expert group that was established by the European Commission in June 2018.

Transparency	Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations
Diversity, Non-discrimination and Fairness	Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle
Societal and Environmental Well-being	AI systems should benefit all human beings, including future generations
Accountability	Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes

From June 2018 to June 2020 the **Assessment List for Trustworthy AI (ALTAI)** benefited from a piloting phase, receiving inputs and feedback from selected companies as well as technical and non-technical stakeholders: during this time, *fifty in-depth interviews with selected companies were conducted thus allowing to collect valuable feedbacks. Moreover, inputs were given through an open work stream on the AI Alliance to provide best practices; and, via two publicly accessible questionnaires for technical and non-technical stakeholders*.

In June 2020 the AI HLEG published the final ALTAI [3].

The ALTAI was intended to protect people's fundamental rights (as enshrined in the EU Treaties, the Charter of Fundamental Rights - the Charter -, and International Human Rights Law). The ALTAI is conceived for self-evaluation purposes. It helps organisations understand what trustworthy AI is, what risks it might generate, what kind of measures may be needed to minimize those risks while maximising the benefits.

Also, that document aims to help organizations to develop an AI application that is lawful, ethical and robust, thus contributing to realize a responsible and sustainable AI innovation in the EU, and enable "responsible competitiveness" [3].

More information regarding ALTAI principles is provided in D4.1 'White Paper for LEAs'.

1.1.3 The EU AI Act Proposal

A Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) was issued by the European Commission (EC) on 21 April 2021. It constitutes a proposal for a Regulation with direct effect on the legal orders of the EU Member States, and follows a human centric and risk-based approach. The AI Act Proposal's scope extends beyond the EU to providers placing on the market or putting into service AI systems in the European Union, irrespective of whether those providers are established within the Union or in a third country as well as to providers and users of AI systems that are located in a third country, where the outputs (i.e., predictions, recommendations or decisions) produced by the AI system are used in the Union.

It is essential to mention that AI systems specifically developed for the sole purpose of scientific research and development are excluded from the scope of the Proposal for a Regulation. However, under all circumstances, any research and development activity should be carried out in accordance

with the Charter on Fundamental Rights and Freedoms of the European Union, EU law as well as the national law and based on the AI HLEG Ethics Guidelines as mentioned in the previous section.

An overview of the latest developments regarding the draft AI Act is presented in D4.1 'White Paper for LEAs'.

1.2 Definition and terminology

This section provides the definitions of the key expressions delineating the work carried out in subtask 5.5 (a), i.e. roadmap, standard, certification, and compliance mechanism for AI regulation. In the appendix, some additional information on standards and compliance mechanism can be found.

- **Roadmap**

This document adopts the following definition: "*A roadmap is a plan that shows how a product or service is likely to develop over the time. […] A roadmap makes it clear what you're trying to achieve and the steps you''ll take towards that end goal*". [4] Such a plan should be easy to understand and should illustrate the actions and the related actors involved needed in order to achieve that specific objective. Roadmaps might include phases with specific timeframe: however, it should be mentioned that roadmaps – especially when addressing a long timeframe – do allow a certain level of flexibility: indeed, as they are supposed to catch the needs and urgency if the future, they might require some adjustments along their implementation in order to fully meet the requirements of the environment.

- **Standard**

Standards are *"technical specifications defining requirements for products, production processes, services or test-methods. These specifications are voluntary. They are developed by industry and market actors following some basic principles such as consensus, openness, transparency and non-discrimination. Standards ensure interoperability and safety, reduce costs and facilitate companies' integration in the value chain and trade"* [5].

The achievement of clear regulation for AI necessitates the establishment of rules and benchmarks. These rules and benchmarks must be incorporated into a comprehensive catalogue of requirements that comprises technical regulations. The purpose of this endeavour is to facilitate the classification of AI systems in accordance with business needs, regulatory policies, and the interests of consumers and end users. Furthermore, these rules and benchmarks are essential in ensuring the reliability, fairness, and transparency of AI technologies. Ultimately, they will establish the fundamental standards that must be met to obtain certifications [6] [7].

- **Certification**

Certification is a formal document issued by independent authority that serves as a guarantee, attesting that a product, service, or system complies with specified requirements. By obtaining certification, providers can validate adherence to high-quality standards, thereby enabling them to design AI applications in a manner that is not only lawful but also ethically acceptable.

A recognized certification, based on ethical and legal requirements, might be the system to create trust on the use of AI.

A certification, based on quality standard, should allow different providers to be compared (that could promote open competition in AI application) and will guarantee the safety and fundamental-rights of people and businesses [6].

- **Compliance mechanism for AI regulation**

In the pursuit of establishing legal frameworks for AI systems several international organizations are currently engaged in the process of drafting or have already reached consensus on high-level rules and ethical guidelines. Such guidelines chiefly encompass five main principles: non-maleficence, autonomy, justice, explicability and beneficence [10].

Compliance mechanisms aimed at ensuring legal conformity throughout the entire lifecycle of the AI systems necessitate a differentiation between ex-ante (i.e. entailing an evaluation of the AI system's adherence to the legal framework before it is deployed on the market) and ex-post assessment of compliance (i.e. involving the monitoring of AI system after they have been introduced on the market). The adaptability of these compliance mechanisms is vital, given the rapidly evolving nature of AI systems [11] [12].

1.3 Structure of the deliverable and relation to other Work Packages

Part A of this document is devoted to the analysis of certification and standards related to artificial intelligence, thus aiming to define a short-term compliance and certification roadmap guiding LEAs and policy makers. As the analysis takes into consideration legislative documents and guidelines, the content is strongly linked to the work carried out within Task 2.2 *Legal framework and casework taxonomy: emerging trends and scenarios on the legislative framework*. Moreover, it is also linked to T3.1 and its deliverable D3.1 *Map of AI in policing innovation ecosystem and stakeholders*, which includes valuable information on the institutional framework and the type of stakeholders that are involved in the AI domain.

The short-term compliance and certification roadmap will provide inputs to Task 1.5 and the second release of the Policy Brief. This document includes recommendations for multiple stakeholders, including policy makers.

This part is structured into four sections.

Section 1 provides an introduction to the topic, describing the background and defining the key expressions that are used throughout the document.

Section 2 presents the methodology adopted to perform the task, including a brief description of the types of activity carried out.

Section 3 reports the results of the desk research activity, illustrating the state of the art regarding certifications, standards and guidelines at national, European and international level. The results of this analysis are included in a catalogue, where standards, certifications and other relevant documents are mapped according to the ALTAI principles.

Section 4 illustrates the compliance and certification roadmap, providing information about the steps, actions and main actors involved.

The conclusions are included in section 10.

2 Methodology

The methodology followed to complete subtask (a) of T5.5 is illustrated in Figure 1 and consists of four main activities, carried out in subsequent phases:

- Desk research
- Creation of an AI Standards catalogue
- Analysis of gaps, overlapping or clashing approaches
- Definition of the roadmap

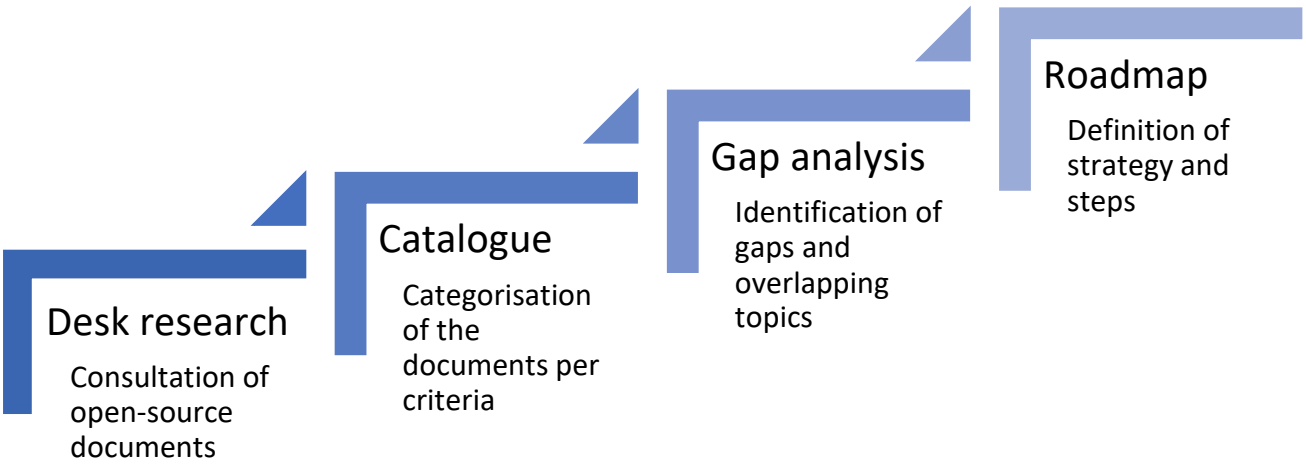

Figure 1 - Methodology for task 5.5 (a)

The first phase consists of desk research in order to gather information on current framework, standards and guidelines, at national, European and international levels. The documents used are open access and include official documents from regulatory authorities, research papers, policy papers, publications from research institutes, national, European or international guidelines on the topic.

Once all these documents have been gathered and analysed, they have been organised in a catalogue (phase 2). The rationale of this phase is to help researchers having a more precise picture of the legislative and regulatory framework, by mapping such documents according to specific criteria. This task is preparatory for the next phase: organising documents according to specific criteria facilitates the identification of commonalities and differences.

The third phase consists of identifying any gaps in the regulations, policies, and guidelines currently existing as well as potential inconsistencies. The intention is to spot those areas where either some clarification should be needed or where there are normative gaps. This activity will help defining the roadmap steps (phase 4) considered as necessary actions to achieve the ultimate goal of having a compliance and certification roadmap.

3 Analysis of the State-of-the-Art of Standardisation and Certification standards

This section includes the results of the desk research carried out in the initial phase of task 5.5 (a). The purpose was to gather a comprehensive understanding of the existing frameworks and approaches in term of standardisation and certification related to AI technologies. The documents analysed have been organised according to different criteria and examined taking into account the ALTAI principles with the purpose of identifying which one(s) is (are) covered by each of these frameworks.

This work will allow to understand which areas are more regulated and which may require stronger regulation, thus being helpful also to inform the policy recommendation preparation (T1.5).

The catalogue (available in Table 2) is organised according to three levels: international, European and national.

The international section includes documents that either address the topic from a global perspective or are drafted by international organisations/entities (could be an SDO but also a private company). The European level embraces sources that are issued by a European entity or authorities and/or apply to the regional landscape. The national level includes documents that contain information on national approaches, whether they are issued by a national authority or not.

For each source, the catalogue indicates the following information:

- Name of the organisation;
- Type of organisation (e.g. research centre, IT company, etc.);
- If the organisation is an SDO, SSO or ESO;
- The type of document (e.g. White paper, research report, article, etc.);
- The year of publication.

Besides the information on the source and its author, the catalogue maps each document into the ALTAI principles, thus indicating which of these principles are addressed or covered by the related document. Additional principles have been added to the table, based on their presence across multiple documents.

Each entry contains the link to the document to facilitate the consultation of the sources.

Table 2 - popAI Standards Catalogue

Information source						ALTAI Principles							Other Principles
Organization	**Type of organization**	**SDO / SSO / ESO**	**Type of document / initiative**	**Year of publication**	**Title**	**Human agency and oversight**	**Technical robustness and safety**	**Privacy and data governance**	**Transparency**	**Diversity, non-discrimination and fairness**	**Environmental and social well-being**	**Accountability**	
International level													
AI Ethics Impact Group	interdisciplinary consortium	No	Research report	2020	From Principles to Practice An interdisciplinary framework to operationalise AI ethics		Reliability	Privacy	Transparency		Environmental sustainability	Accountability	Justice
Association for Talent Development	non-profit association	No	Article	2023	7 Principles to Guide the Ethics of Artificial Intelligence		Security Reliability	Privacy	Transparency	Fairness Inclusiveness		Accountability	
ATP Global	dissemination company	No	Blog / Publication	2022	Artificial Intelligence Principles	Human-in-the-loop Balanced utilization			Transparency	Fair and unbiased			Balances utilisation
Ceweb	IT company	No	Website	2020?	MAPPING PRINCIPLES OF ARTIFICIAL INTELLIGENCE		Reliability and safety	Privacy and security	Transparency	Fairness	Social impact	Accountability	

Chambers and partners	legal research company	No	Research report	2022	Governance Guidelines for Implementation of AI Principles Ver. 1.1	Human-centric principle	Principle of ensuring security	Principle of privacy protection	Principles of fairness, accountability and transparency	Principle of fair competition Principles of fairness, accountability and transparency		Principles of fairness, accountability and transparency	Principle of education and literacy Principle of innovation
Cisco	IT company	No	Report	2022	Cisco Principles for Responsible Artificial Intelligence		Security Reliability	Privacy	Transparency	Fairness		Accountability	
Committee for economic development of Australia	non-profit organisation	No	Paper	2020	AI PRINCIPLES TO PRACTICE	Human-centered values Contestability	Reliability and safety Privacy protection and security	Privacy protection and security	Contestability Transparency & explainability	Contestability Fairness	Human, societal & environmental wellbeing	Contestability Accountability	
Dell	IT company	No	Internal guidelines	N/a	Dell Technologies Principles for Ethical Artificial Intelligence	Responsible			Transparent	Equitable	Beneficial	Accountable	
Digital Dubai	digital economy company	No	Internal guidelines	N/a	Principles of artificial intelligence	Ethics	Security	Security	Humanity	Inclusivness	Humanity		

ETSI	not-for-profit institute	ESO	Research report	2020	Artificial Intelligence and future directions for ETSI		Network optimization and end-to-end assurance	IoT, data acquisition & management, governance and provenance Security and privacy			Health and societal application of AI		Testing
Forrester	research and consultancy company	No	Report	2020	Five AI Principles To Put In Practice		Privacy and security	Privacy and security	Trust and transparency	Fairness and bias	Social benefit	Accountability	
Future of Life Institute	research institute involved into artificial intelligence, nuclear technology and biotechnologies	No	Report	2017	AI Principles	Responsability Human values	Safety Failure transparency	Safety Personal privacy Liberty and privacy	Failure transparency Judicial transparency	Shared prosperity		Responsability Human control	Cost and benefit Liberty and privacy Judicial transparency and responsibility
Google	internet services	No	Internal guidelines	N/a	Objectives for AI applications		Be built and tested for safety	Incorporate privacy design principles		Avoid creating or reinforcing unfair bias	Be socially beneficial	Be accountable to people	Uphold high standards of scientific excellence

Hewlett Packard Enterprise	IT company	No	Internal guidelines	N/a	HPE AI ETHICS AND PRINCIPLES	Human focused principle	Robust principle	Privacy-enabled security		Inclusivity principle			
IBM	IT company	No	Internal guidelines	N/a	AI Ethics		Robustness	Privacy	Explainability Transparency	Fairness			
IEEE Standards Association	standards association	SDO	Research report	2018	ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems	Human rights			Transparency	Human rights A/IS technology misuse and awareness of it	Prioritizing well-being	Accountability A/IS technology misuse and awareness of it	
IEEE Standards Association	standards association	SDO	Report	2022	Trustworthy AI	Verifiability Autonomy and control	Safety Robustness and reliability	Privacy Security	Transparency Explainability	Non-discrimination, bias and fairness	Sustainability	Verifiability	Functionality and performance
IEEE Standards Association	standards association	SDO	Research report	2020	IEEE USE CASE—CRITERIA FOR ADDRESSING ETHICAL CHALLENGES IN TRANSPARENCY, ACCOUNTABILITY, AND PRIVACY OF CONTACT TRACING—DRAFT CALL FOR GLOBAL CONSULTATION			Privacy	Transparency			Accountability	

IEEE Standards Association	standards association	SDO	Report	N/a	Verifying Ethics in Ai-based solutions			Privacy	Transparency	Algorithmic bias		Accountability	
IFPMA	pharmaceutical international organization	No	Report	2022	IFPMA Artificial Intelligence Principles	Empowering humans Human control	Privacy, security and safety by design	Privacy, security and safety by design	Transparency, explainability and ethical use	Fairness and minimization of bias		Accountability	Transparency, explainability and ethical use
International Research Center for AI Ethics and Governance	research center	No	Website	2019	A Cross Cultural and Transdisciplinary Center for Building Responsible andBeneficial AI for Human and Ecology Good	Be Responsible Be Ethical				Be Diverse and Inclusive	Do GoodFor Humanity	Be Responsible	Cost and benefitControl risks
Joint Artificial Intelligence Center	Intelligence government organization	No	Internal guidelines	2020	Ethical Principles for Artificial Intelligence	Responsible	Reliable			Equitable		Traceable Governable	

Massachusetts Institute of Technology	university	No	Report	2020	MIT Technology review	Public trust Public participation	Scientific integrity and information quality Safety and security	Safety and security	Disclosure and transparency	Fairness and non-discrimination			Cost and benefits Flexibility Risk assessment and management Interagency coordination
McKinsey	consultancy company	No	Website	2022	Ethical AI Principles	Human oversight and accountability	Performance and safety Security	Privacy and data ethics Security	Explainability and transparency	Bias and fairness	Sustainability	Human oversight and accountability	Performance and safety
National Association of Insurance Commissioners	standard-setting organization	SSO	Report	2020	National Association of Insurance Commissioners (NAIC) Principles on Artificial Intelligence (AI)	Fair and ethical	Secure, safe and robust	Secure, safe and robust	Transparent	Fair and ethical		Accountable	Compliancy to performance standards
NATO	international organization	No	Website	2021	Summary of the NATO Artificial Intelligence Strategy	Governability	Reliability		Explainability and traceability	Bias mitigation		Responsability and accountability	Lawfulness
Philips	electronic and healthcare	No	Internal guidelines	N/a	Philips AI Principles	Oversight	Robustness		Transparency	Fairness	Well-being		

	company												
Prolific	research center	No	Website	N/a	What are AI ethics? 5 principles explained		Reliability Security and privacy	Security and privacy	Transparency	Impartiality		Accountability	
PWC Australia	consultancy and legal company	No	Website	2022	Ten principles for ethical AI	Human agency	Reliability and robustness Safety	Privacy Security		Fairness	Beneficiality	Accountability	Interpretability Lawfulness
SAP	IT company	No	Website	N/a	SAP's Guiding Principles for Artificial Intelligence	Design for people	Transparency and integrity Quality and safety standards	Data protection and privacy	Transparency and integrity	Business beyond bias	Societal challenges		Quality and safety standards
Spark	telecommunication company	No	Report	2022	Spark's Artificial Intelligence Principles	Ethical design Informed human decision making		Privacy	Transparency and explicability	Diversity, inclusivity and bias		Informed human decision making	
Telefonica	telecommunication company	No	Report	2018	AI Principles of Telefonica	Human-centric	Privacy and security by design	Privacy and security by design	Transparent and explainable	Fair			

Telia Company	Technology company	No	Website	N/a	AI ethics	Responsible and value centric Human centric	Safe and secure	Safe and secure	Transparent and explainable	Fair and equal		Responsible and value centric Accountable Control	Continuous review and dialogue Respect of rights
Xenonstack	IT company	No	Website	N/a	Responsible AI Principles and Challenges for Businesses	Human augmentation	Data risk awareness	Trust by privacy Data risk awareness	Explainability by justification Reproducible operations	Bias evaluation	Human augmentation		Displacement strategy Practical accuracy
The Institute for Ethical AI & Machine Learning	Research centre	No	Website	N/a	The Responsible Machine Learning Principles	Human augmentation	Data risk awareness	Trust by privacy Data risk awareness	Explainability and justification	Bias evaluation		Displacement strategy	Reproducible operations Practical accuracy Data risk awareness
The Royal Australian and New Zealand College of	international organization	No	Report	2019	Ethical Principles for Artificial Intelligence in Medicine	Application of human values Governance	Safety	Privacy and protection of data	Transparency and explainability	Avoidance of bias		Responsibility for decisions made	Decision-making on diagnosis and treatment Teamwork Responsibil

Radiologist													ity and accountability
The Wolfsberg Group	Bank association	No	Report	2022	Wolfsberg Principles for Using Artificial Intelligence and Machine Learning in Financial Crime Compliance				Openness and transparency			Accountability and oversight	Legitimate purpose Proportionate use Design and technical expertise
Thomson Reuters	mass media and information company	No	Website	N/a	Artificial Intelligence Our AI principles	Human-centric approach	Safety Security Reliability	Privacy		Social benefits		Accountability	
Turkcell	service provider	No	Website	N/a	Turkcell Artificial Intelligence Principles	Human and environment centric	Security-based	Data privacy Security-based	Transparent	Fair	Human and environment centric Collaboration for a better future	Responsible	
U-next	learning company	No	Website	2022	What Are Explainable Artificial Intelligence (AI) Principles?	Human-centered design			Explainability Transparency				Empathy

European level													
CENTRIC and Europol (AP4AI project)	intelligence companies	No	Research paper	2022	Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain	CompellabilityEnforceabilityandRedressConductLearning organization	Commitment toevidence		Transparency Explainability	PluralismConstructivness			UniversalityIndependenceLegality
Fraunhofer Institute	science research not-for-profit institute	No	Research paper	N/a	Trustworthy use of artificial intelligence		Security Reliability	Data protection	Transparency	Autonomy and control		Fairness	Ethics and law
Fraunhofer Institute	science research not-for-profit institute	No	Research paper	2021	Ai assessment catalog	Trustworthiness	Safety and security Reliability	Data protection	Transparency	Fairness		Autonomy and control	
European Commission	regional organisation	No	White paper	2020	White Paper on Artificial Intelligence: a European approach to excellence and trust	Human agency and oversight	Technical robustness and safety	Privacy and data governance	Transparency	Diversity, non-discrimination and fairness	Environmental and social well-being	Accountability	
Joint Research Centre	research centre	No	Technical report	2022	AI Watch: European Landscape on the Use of Artificial Intelligence by the Public Sector	Stimulating awareness and knowledge sharing	AI infrastructure						Testing Improving data access and quality Improving internal capacity

Joint Research Centre	research centre	No	Technical report	2023	AI Watch: Artificial Intelligence Standardisation Landscape Update	Human oversight	Accuracy, Robustness and Cybersecurity	Data and data governance Record keeping	Transparency and Provision of Information to users			Quality Management System	Record keeping and risk management
OECD.AI Policy Observatory	policy observatory	SSO	Website	2019	Standardisation Landscape Update	Human-centered values and fairness	Robustness, security and safety	Robustness, security and safety	Transparency and explainability	Human-centered values and fairness	Inclusive growth, sustainable development and well-being	Accountability	
National level													
United Nations	international organization	No	Report	2022	Principles for the ethical use of artificial intelligence in the United Nations system	Human autonomy and oversight	Safety and security	Safety and security Right to privacy, data protection and data governance	Transparency and explainability	Fairness and non-discrimination	Sustainability	Responsibility and accountability	Defined purpose, necessity and proportionality Do no harm
Australia government	government	No	Website	N/a	Australia's AI Ethics Principles	Human-centred values	Reliability and safety	Privacy protection and security	Transparency and explainability	Fairness Contestability	Human, societal and environmental well-being	Accountability	

Australian New South Wales regional government	government	No	Website	N/a	Mandatory Ethical Principles for the use of AI		Privacy and security	Privacy and security	Transparency	Fairness		Accountability	
Chambers and partners	legal research company	No	Guidelines	2022	Governance Guidelines for Implementation of AI Principles Ver. 1.1	Human-centric principle	Principle of ensuring security	Principle of privacy protection	Principles of fairness, accountability and transparency	Principle of fair competition Principles of fairness, accountability and transparency		Principles of fairness, accountability and transparency	Principle of education and literacy Principle of innovation
DIN e. V. and DKE	private companies	No	Research report	2020	German Standardisation Roadmap on Artificial Intelligence		IT security (and safety) in AI systems	IT security (and safety) in AI systems		Ethics/Responsible AI			Quality, conformity assessment and certification
Joint Research Centre	research centre	No	Technical report	2022	AI Watch: Defining Artificial Intelligence	Perception			Communication Learning	Ethics and Philosophy			Services Integration and Interaction Learning, planning and reasoning

Joint Research Centre	Research centre	No	Technical report	2022	AI Watch National Strategies on Artificial Intelligence: A European Perspective					Inclusion and diversity			Upskilling in three main sectors: at-risk professional profiles, teachers and public servants Assessment of effectiveness and impact of skilling measures
Malta Digital Innovation Authority	government authority	No	Guidelines	2019	AI Innovative Technology Arrangement Guidelines		Integrity		Transparency			Accountability	Compliancy to performance standards Legality
Agenzia per l'Italia digitale	government authority	No	Guidelines	2018	Libro Bianco sull'intelligenza artificiale (The White book on Artificial Intelligence)		Securing access to data and computing infrastructures	Securing access to data and computing infrastructures			Promoting the adoption of AI by the public sector	Partnership with the private sector International aspects	Testing Cost and benefit

National Institute of Standards and Technology	government agency	No	Research report	2019	U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools	Societal and ethical considerations Human-centered		Privacy		Societal and ethical considerations		Governance	Innovation-oriented Applicable across sectors
US Department of Defense	government-related department	No	Website	2020	DOD Adopts Ethical Principles for Artificial Intelligence	Governable	Reliable		Traceable	Equitable		Responsible	
US Office of the Director of National Intelligence	Intelligence government organization	No	Guidelines	N/a	Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf (dni.gov)	Human-Centered Development and Use	Secure and Resilient	Secure and Resilient	Transparent and Accountable	Objective and Equitable		Transparent and Accountable	Respect the Law and Act with Integrity Informed by Science and Technology

4 Definition of the roadmap

The compliance and certification roadmap defined within Task 5.5 (a) is outlined in the context of the approval of the EU AI Act and, thus, aims to inform the EC on specific aspects related to the use of AI-based technologies by LEAs.

The roadmap stems from the analysis of existing gaps and overlapping provisions found out during the desk research phase and intersects such considerations with the need to ensure the implementation of the AI Act across the EU. The mapping of research and policy papers, guidelines, and legislative framework helped to point out "grey areas" i.e. aspects where guidelines or regulatory frameworks are missing or do not provide a common and unique approach. In parallel, the analysis of the EU AI Act provides a comprehensive view of the current status at EU level it has not entered into force yet.

This roadmap, hence, aims to suggests a course of action that should support the implementation of the AI Act as well as a common and unique framework for the certification of AI technologies in Law Enforcement.

The roadmap includes specific actions to undertake and address multiple stakeholders, ranging from policy makers to technology providers.

This roadmap is intended as a flexible approach that could be adjusted according to the future evolution of the context.

The short-term roadmap presented in the first part of this deliverable is illustrated in Figure 2 and structured in 6 steps, described more in detail in Table 5.

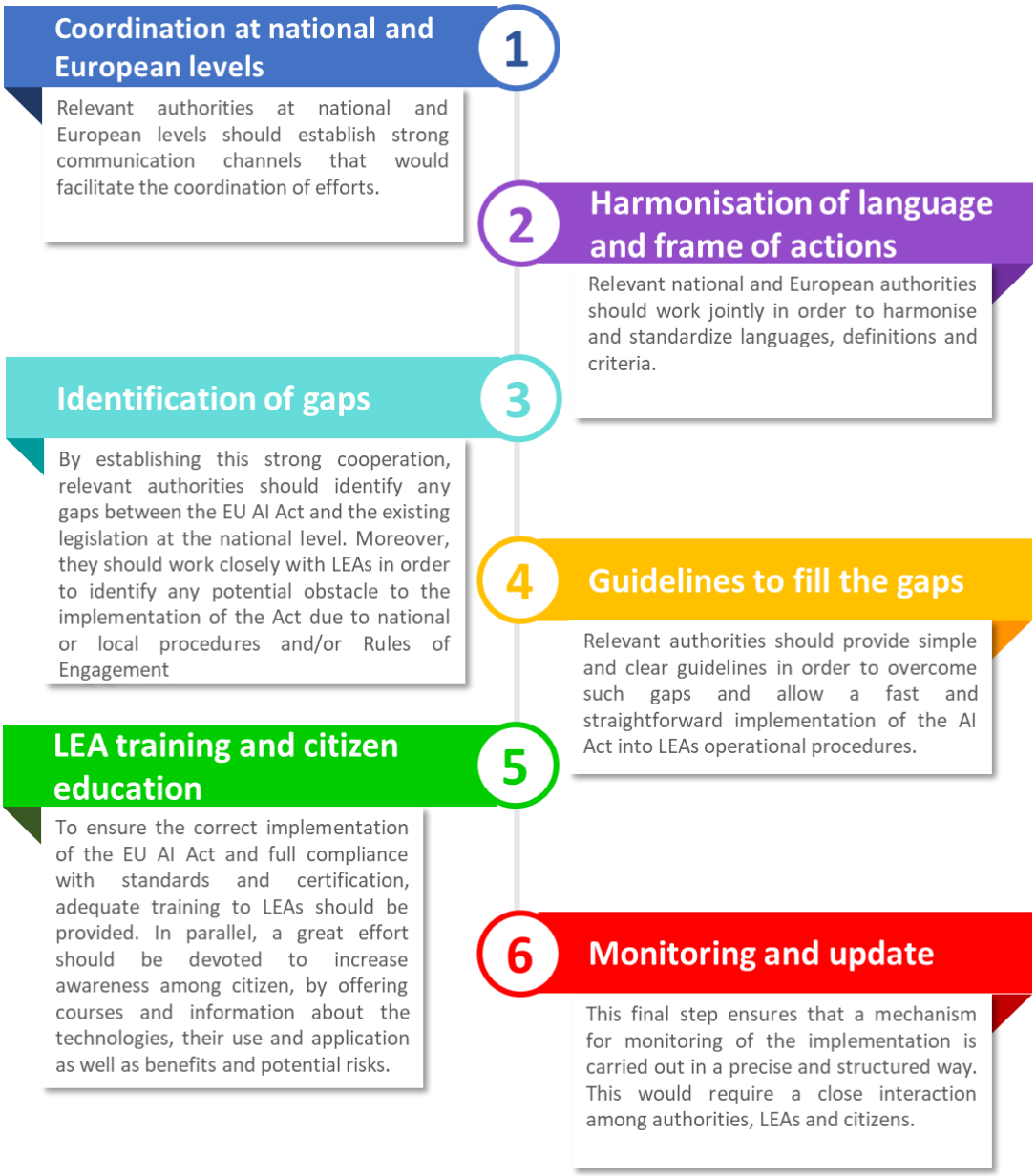

Figure 2 - popAI certification and compliance roadmap

Table 3 - popAI certification and compliance roadmap (steps)

Step 1	Coordination at national and European Levels
Description of the action	Relevant authorities at national and European levels should establish strong communication channels that would facilitate the coordination of efforts. As the EU AI Act stands as the first AI legal framework at European level aiming to regulate the use and application of AI-based technologies, it is crucial that all relevant authorities that deal with the standardisation, the certification and the definition of operating procedures work jointly to ensure the adequate application of such principles. The coordination among these actors should allow a prompt and effective implementation of the provisions of the Act, as well as the compliance of the different actors to the new framework.
Targeted stakeholder	Regulatory authorities; LEAs; certification and standardisation organisations

Step 2	Harmonisation of language and frame of action
Description of the action	Relevant national and European authorities should work jointly in order to harmonise and standardize languages, definitions and criteria. One of the key aspects when regulating a field are definitions. With regard to artificial intelligence, the issue of setting boundaries and defining criteria it has been quite critical, even prior to the preparation of the EU AI Act. This harmonisation of language and types of actions across European countries is critical for a correct and efficient implementation of any regulatory framework. Such definitions should be set clearly and accepted in order to avoid confusion or misunderstanding.
Targeted stakeholder	Regulatory authorities; certification and standardisation organisations; industry; LEAs

Step 3	Identification of gaps
Description of the action	By establishing this strong cooperation, relevant authorities should identify any gaps between the EU AI Act and the existing legislation at national level. Moreover, they should work closely with LEAs in order to identify any potential obstacle to the implementation of the Act due to national or local procedures and / or Rules of Engagement.
Targeted stakeholder	Regulatory authorities; LEAs; certification and standardisation organisations

Step 4	Guidelines to fill gaps
Description of the action	Relevant authorities should provide simple and clear guidelines in order to fill such gaps and allow a fast and straightforward implementation of the AI Act into LEAs operational procedures.
Targeted stakeholder	Regulatory authorities and LEAs

Step 5	LEA training and citizen education
Description of the action	To ensure the correct implementation of the EU AI Act and full compliance with standards and certification, adequate training to LEAs should be provided. In parallel, a significant effort should be devoted to increase awareness among citizen, by offering courses and information about the technologies, their use and application as well as benefits and potential risks. Citizen education would ensure understanding of actions to undertake, as well as ensure transparency in what technologies will be used for, their limitations, benefits, risks and mitigation measures. This would increase people's awareness and perhaps facilitate a smoother acceptance of the technology in the security domain.
Targeted stakeholder	Regulatory authorities; citizens; LEAs; industry

Step 6	Monitoring and update
Description of the action	This final step ensures that a mechanism for monitoring the implementation is carried out in a precise and structured way. This would require a close interaction among authorities, LEAs and citizens. Specific fora should be established where all interested parties could illustrate needs and problems, thus permitting to overcome potential challenges and arrange adequate actions when and where needed.
Targeted stakeholder	Regulatory authorities; LEAs; certification and standardisation entities; citizens; industry

This roadmap should be received as a set of consequential actions to be implemented in the short term, in order to support the effective implementation of the EU AI Act. It includes activities and recommendations to different stakeholders, thus recognising the importance of a joint and commonly accepted action. Transparency and cooperation are of utmost importance to fulfil this goal, and a close monitoring of the actions would avoid diversion and ensure the correct implementation of the roadmap.

It should also be noted that a constant interaction among the diverse stakeholders would allow making all necessary adjustments to the roadmap, should new needs come up requiring immediate actions.

The collaboration among the multiple actors that have interests at stake when it comes to artificial intelligence technologies and their application into the security domain is the backbone of popAI project. Opportunities for exchange of knowledge and experiences, as well as challenges and lessons learnt have been created during the project implementation showing the importance of this mutual understanding and the beneficial value of information sharing to improve the coordination of the efforts. This short-term roadmap traces this guidance, thus suggesting a close interaction as well as a combination of bottom-up and top-down approaches, encompassing directives from the authorities as well as inputs from the society.

Part B – Subtask (b) – popAI Roadmap for 2040

5 Introduction

Part B encapsulates the results of the research conducted under subtask (b), aiming to define a long-term roadmap highlighting "*potential scenarios and futures, risks and strategies to get there*". The roadmap looks into 2040: a set of future scenarios have been defined, stemming from the results of the foresight scenarios activity carried out in WP3. The purpose is to foresee how the future could evolve, thus including new challenges, potential risks and needs.

Once these elements have been identified, the roadmap indicates the actions needed to address new threats and risks, bearing in mind also the increasing development and enhancement of AI-based technologies and the related risks but also opportunities they bring along.

As mentioned for the short-term roadmap, the popAI 2040 roadmap is meant to be a guideline for future actions that may be adjusted should new needs emerge. It represents "*a policy and practice-oriented resource that will capture the contributions of the EU towards building better but also responsible, ethical and value-based AI tools for LEA use*".

The roadmap is articulated into three strands, covering the key aspects related to AI-base systems and their employment in the security sector. The technological strand indicates actions that regard the development and/or testing of the future technologies. The organisational strand suggests adjustments that should facilitate the introduction of AI systems into operational procedures, while providing common frameworks and actions. The regulatory strand indicates relevant actions to undertake in order to ensure that the legislative framework keeps pace with the evolution of the technologies as well as the emergence of new societal needs.

The 2040 roadmap provides also input to the second policy brief (Task 1.5).

5.1 Background

5.1.1 The EU AI Strategy

As mentioned, task 5.5 (b) looks into the future and aims at defining actions to ensure that a future application of AI technologies in security tasks would be legal and ethical compliance. However, when planning for the future, it is crucial to know current activities and initiatives, to ensure the proposed measures are compatible and sustainable.

This section, hence, provides information of the initiatives promoted at the EU level to deal with the increasing role of AI in the society. The EU has been addressing proactively the issue of AI and its impact on society. The EU approach to artificial intelligence gravitates around the concept of excellence and trust, to be considered as the foundation of research and industrial capacity development.

The European AI Strategy aims to make the EU a world leader in AI, ensuring trustworthiness and protection of the fundamental values of the Union and the respect of human rights. Such an ambition has led to multiple initiatives and actions aiming to set concrete rules for the development and use of AI.

Figure 3 indicates the core milestones towards the adoption of the so-called EU AI Act, aimed to be the world's first comprehensive regulatory framework for artificial intelligence.

The Commission aims to address the risks generated by specific uses of AI through a set of complementary, proportionate and flexible rules. These rules will also provide Europe with a leading role in setting the global gold standard. The legal framework for AI proposes a clear, easy to understand approach, based on four different levels of risk: unacceptable risk, high risk, limited risk, and minimal risk [19].

June 2023 has been a critical moment for the implementation of the EU AI Act.

Furthermore, during the last State of the Union Address by EC President Ursula Von der Leyen reiterated the fast and growing importance of artificial intelligence in multiple societal sectors, as well as the need to understand the potential threats deriving from it. Indeed, although AI offers several windows of opportunity, it also hides some risks. A regulatory framework should help to exploit the benefits of this technology while avoiding or minimising the possible risks.

Figure 3 - Milestones towards the EU AI Act

"I believe Europe, together with partners, should lead the way on a new global framework for AI, built on three pillars: guardrails, governance and guiding innovation. ", mentioned President Von der Leyen. Guardrails refer first and foremost the EU AI Act, as a blueprint for the entire world. Governance recalls the idea of a single governance system in Europe, where the different partners should join forces to ensure a global approach to understanding the impact of AI in our societies. Lastly, guiding innovation refers to the need for an open dialogue with those that develop and deploy AI [20].

popAI has conducted its activities in a critical moment, thus having the opportunity to be part of this effort towards the definition of a regulatory framework that could support an ethical and trustworthy use of AI at a global level. Its inclusive approach would permit to bring the voices from the multiple categories of stakeholders to the policy making level, thus contributing to ensure that provisions and rules manage to meet the needs of the broad society.

5.2 Definitions and terminology

- **Foresight scenario**

The topic of foresight scenarios has been thoroughly addressed in deliverable D3.5 *Foresight scenarios for AI in policing.* Hence, a full and detailed description of the concept, the main theories and the approaches to scenario building can be found in that document.

- **Roadmap**

The definition of roadmap has already been presented in part A of this deliverable. Please refer to section 1.2.

- **Strategy**

The word *strategy* does not have a common and unique definition although is generally associated to the achievement of a final goal and it is about maintaining a balance between three factors: end, ways and means. The ends are the final goal that someone wants to achieve, the means are the resources available to reach such ends and, lastly, ways are the modalities the means are employed to reach the end [16]. When considering strategy, this deliverable refers to the definition and analysis provided by Lawrence Freedman in his work "Strategy", who investigates thoroughly this topic. According to Freedman, strategy is often described as a sequence of actions towards a desired end state, thus gravitating around this final goal. However, the expert suggests that reality is quite different. The process of strategy "*evolves through a series of states, each one not quite what was anticipated or hoped for, requiring a reappraisal and modification of the original strategy, including objectives*". Strategy is, hence, understood as a flexible and fluid approach, built around the starting point and not the end state [16].

In the framework of popAI, this concept of strategy sounds more adequate. Indeed, given the extreme speed of the technology evolution and the constant progresses in the ethical and legal frameworks, it is fundamental to adopt a flexible approach, an emergent strategy that could be able to keep pace with and adapt to the evolving landscape, thus being truly able to meet the existing needs.

5.3 Structure of the deliverable and relation to other WPs

The activities carried out within WP3 have been of utmost importance for the definition of the roadmap. The policy labs have offered the opportunity to understand LEAs' perspective in terms of current and potential future challenges and needs. Their voice is a vital element in order to understand how AI could actually support their activities now and in the future. The foresight

scenarios have been used as a starting point to picture the possible evolution of the security contest in a 20-years perspective.

popAI WP4 provides recommendations for the ethical use of AI by LEAs, which are both derived and addressed to stakeholders, including LEAs, policymakers, the civil society and technology developers. The recommendations provide insights to the needs and concerns identified under the popAI project and emerging best practices on the use of AI by LEAs according to the applicable legal framework and ethical principles. They also map the latest developments regarding the AI Act Proposal and aspire to be complementary to the forthcoming legal framework to be regulating AI in the EU. Lastly, they are concerned with some key issues of the AI Act Proposal and are introducing respective suggestions. These recommendations have been taken into account in formulating the 2040 roadmap.

The popAI roadmap to 2040 provides also input to the second policy brief that will be outlined within the framework of Task 1.5. The policy brief is an in-depth evidence-based analysis that addresses the main concerns, risks and threats involved by using and developing AI tools in the security domain, while including practical recommendations for the policy making level.

Part B of this deliverable includes the outcomes of this subtask.

Section 5 provides an introduction to the topic, thus including background information and explaining the terminology.

Section 6 outlines the methodology adopted to carry out the action.

Section 7 reports the outcomes of the literature review carried out to gain a more comprehensive understanding of the security landscape and the current European strategic approach to tackle challenges and threats. This information is a relevant starting point to look into the future and understand how threats could evolve and technology could support actions of Law Enforcement agencies.

Section 8 recalls the findings of WP3 and WP4 that have been used for the purpose of task 5.5. More detailed information about the activities carried out under these WPs can be found in the respective deliverables.

Section 9 outlines the 2040 roadmap, divided into three strands (technological, organisational, and regulatory). For each of them, a description of activities is provided, as well as the target audience.

The conclusions are outlined in section 10.

6 Methodology

The methodology followed to complete subtask (b) of T5.5 is illustrated in Figure 4 and consists of four main activities:

- Desk research (phase 1);
- Collection of input from WP3 and WP4 (phase 2);
- Definition of the roadmap (phase 3);
- Validation of the roadmap (phase 4).

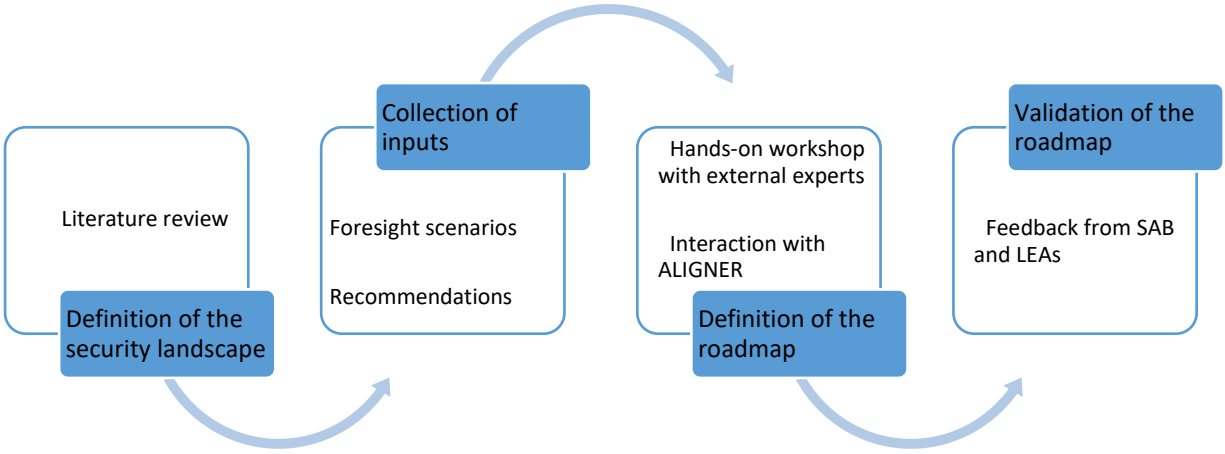

Figure 4 - Methodology for task 5.5 (b)

The desk research aimed to gain a more comprehensive understanding of the current planning at European level in terms of addressing security threats and exploiting the benefits of technological innovation. Indeed, when investigating the future of AI technologies in the security domain it is quite critical to know how the current environment could evolve. Hence, it is crucial to start from the current landscape and understand the key driving factors and trends. What are the major security concerns at the present moment? What are the EU priorities? What are countries doing to face the existing security challenges? This preliminary research is of utmost importance to investigate how these threats and challenges could unfold in the future and, consequently, try to understand how AI technologies could support in tackling them.

The second activity consists of gathering the inputs from WP3 and WP4. In particular, Task 5.5 (b) receives the foresight scenarios developed within Task 3.5, which has exploited the work carried out in the policy labs of Task 3.4. WP4, instead, contributes to the final outcome of Task 5.5 (b) with its pandect of recommendations, which includes three sets of recommendations: to and from LEAs and policy makers; to and from citizens; to and from technology developers.

The third step consists in the definition of the roadmap: in this light, a workshop was conducted with the Stakeholder Advisory Board (section 5.1). In addition, consultations with sibling project ALIGNER were conducted. Given the interactions among the projects and the complementarity of their research, it was considered worth to exchange views and current status of each own research. These two activities helped popAI consortium to draft the roadmap, presented to the Stakeholder Advisory Board and the LEAs within the consortium for validation. The roadmap was also illustrated during popAI final event in Brussels.

7 The European security landscape

This section aims to provide an overview of the European security landscape, thus identifying the current main challenges and threats as well as key strategic approaches. The information included in this section stems from the analysis of multiple documents, including official documents issued by the European Commission but also research and policy papers.

Table 4 lists the key documents that have been taken into consideration.

Table 4 - Main sources used for the desk research

Author	Title	Date of publication
Council of the European Union	European Union Global Strategy	2016
European Commission	EU Security Union Strategy	2020
Council of the European Union	EU Council Conclusion on Security and defence	2021
European Union	EU Strategic Compass	2022
World Economic Forum	Global Risks Report 2023	2023
European Commission	Roadmap on security and defence technologies	2021
European Parliament	Report on Critical technologies for security and defence: state of play and future challenges	2023

The analysis pointed out that since the 1990s, the increasing international instability and security challenges have drawn governments' attention towards the need for enhanced security and defence measures. Security threats have multiplied and diversified in the past decades, thus challenging countries and their ability to adequately deal with them.

The concept of **hybrid war and threats** has become predominant, pointing out the different nature of such threats. Security threats now range from political to economic, from military to social ones. Furthermore, new actors have appeared in the arena, daring the predominant role of nations and contributing to increase the complexity of the international system.

Threats are variegated and range from terrorism and organised crime, to human trafficking and drugs trade. Moreover, the evolution of technology has paved the path to new types of threats such as cyberattacks and cybercrime but also the proliferation of fake news and disinformation, with all the consequences on the political landscape and social composition.

Looking at the international landscape it emerges a very complex picture. The 2023 Davos report illustrates the interconnection of the multiple risks that characterised the international environment: the diagram in Figure 5 below shows such interconnectivity, while highlighting the variegated nature

of these risks and the impact they may have at diverse levels (e.g. societal, economic, geopolitical, etc.) [17].

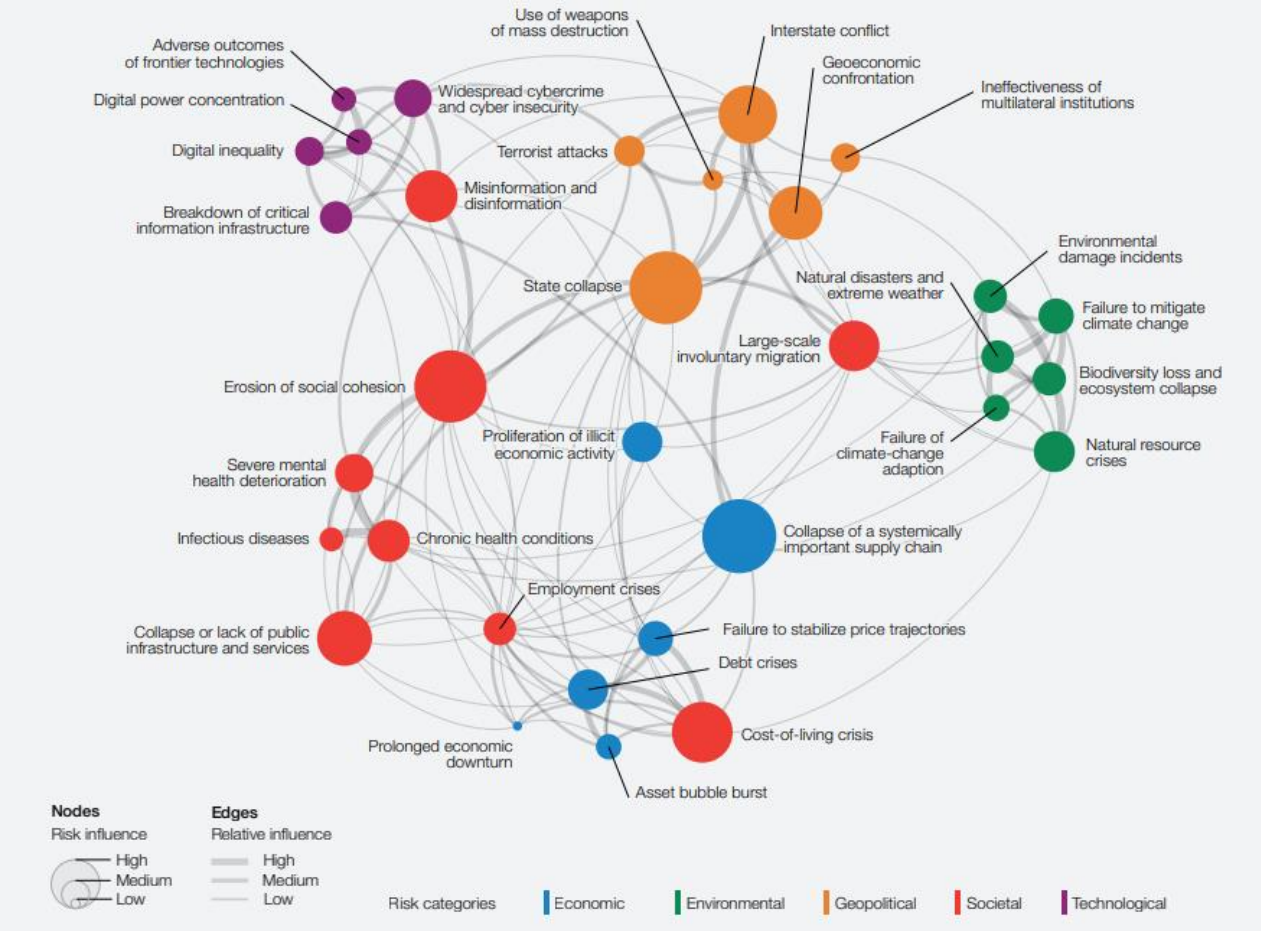

Figure 5 - Global risks landscape: an interconnections map (Source: [17])

As stated in the EU Security Union Strategy, "*Europeans today face a **security landscape in flux**, impacted by evolving threats as well as other factors including climate change, demographic trends and political instability beyond our borders."* Amongst the main security threats, the document lists cybersecurity, cybercrime, hybrid attacks, terrorist attacks and organised crime, which entail multiple types of actors and actions, as well as targets and technologies adopted.

In this context, the EU Security Union Strategy lists 4 priorities that are summarised in Figure 6 below.

Figure 6 - EU Security Union Strategy: priorities (Source: [18])

In this landscape, it is clear how technologies could play a significant role, both in terms of posing new threats and in offering opportunities to improve the management of such threats. As stated in the strategy, law enforcement and justice practitioners need to adapt to the new technologies, both considering them as a potential threat (should they be used with malicious intents) or as a source to facilitate the implementation of their tasks. In addition, the rise of dual-use applications has led to an increased interrelation between security and defence, also in terms of capabilities and assets [18].

It is clear, hence, that LEAs are facing multiple challenges both in terms of threats and operational procedures: the deployment and use of new technologies, indeed, could require some changes in terms of internal processes, organisation of tasks and procedures. Moreover, in view of stronger collaboration and exchange of information across EU Member States, it is quite critical to achieve common understanding of threats as well as management plans.

By bringing together LEAs from five different EU countries, popAI project has managed to contribute to this scope: LEAs have had the opportunity to explain their challenges and constraints as well as exchange views and lessons learnt with colleagues from other countries. During the policy labs and the interactive workshops conducted within the framework of WP3 and WP5, multiple perspectives have emerged, highlighting common challenges as well as specific concerns.

The roadmap presented in section 9 emerges also from the interactions with LEAs and the analysis of what are perceived as the most likely risks in the future and the understanding of what might be needed to ensure an effective fulfilment of LEAs' duties.

8 Inputs from WP3 and WP4

8.1 Input from WP3 foresight scenarios

As mentioned, Task 5.5 (b) takes into account the foresight scenarios developed within Task 3.5 and uses them as a support for the definition of the roadmap. As it will be further explained in section 5.1, the foresight scenarios from Task 3.5 covered a five-year framework. These scenarios have been used to look into a more distant future, trying to foresee how the different situations and challenges addressed in each scenario could evolve in a 20-year timeframe.

The full description of the scenarios is available in Annex III.

In this section, the five scenarios from Task 3.5 are described. The same information was provided to the participants to the workshop.

8.2 Input from WP4 Recommendations for the ethical use of AI by LEAs

The recommendations from WP4 regarding the ethical use of AI by LEAs derived from WP2 framework and WP3 empirical research, and are categorised based on the stakeholders to which they are addressed. WP4 provides, indeed, three sets of recommendations direct to (i) LEAs and policymakers, (ii) civil society and (iii) technology developers. A summary of these recommendations is available in Annex IV, while the full text is included respectively in D4.1, D4.2, and D4.3.

9 Definition of the roadmap

9.1 Workshop with Stakeholder Advisory Board

9.1.1 Purpose and structure of the workshop

The aim of the workshop was to co-develop foresight scenarios depicting the use of AI in policing covering a 20-year time frame to support the definition of the popAI Roadmap for 2040.

These scenarios have been used to investigate how security needs might change in the future and what would be needed to address them. How AI-based technologies could be used in the future in support of security tasks? Would new regulations be necessary in order to ensure a useful and ethical use of such technologies? Would the organisational framework be impacted as well?

This exercise helped to understand to what extent the use of such technologies would meet the needs of the specific security challenges but also to investigate if additional no-technological elements would be required.

The scenarios have, then, been used to define the popAI Roadmap to 2040. This roadmap takes into consideration the future perspective, not just referring to possible future security scenarios but analysing the specific needs that might arise and the means that would be helpful to address them. The roadmap considers actions and measures that could help meet the identified security challenges, considering technological, organisational and regulatory perspectives. The strategy includes information on the involved actors, the actions to undertake, the technologies needed and all those elements that would be relevant to achieve an ethical and lawful use and application of the AI technologies in civil security domain.

The foresight scenarios defined within WP3 were used as a starting point for this exercise. The five scenarios address different threats and situations and cover a time frame of 5 years. They provide a narrative of the situation and describe a potential outcome.

The purpose of this workshop was to further elaborate these scenarios, looking into a 20-years perspective and trying to foresee possible future development and identify new potential needs for LEAs and citizens. Once this future context is defined, it will be discussed which type of changes could be needed in order to address the new security challenges: new technologies? New regulations? Adjustments in the organisational aspects?

Execution

The audience was divided into five groups, ensuring different expertise for each group. Each group was assigned a moderator and the discussion was driven following three steps.

Step 1: each group was illustrated one of the foresight scenarios from WP3. Each scenario addresses a specific security challenge (e.g. border control, migration, crime prevention, etc.) and involves some actors and technologies. The scenarios that emerged from WP3 cover a 5 years' time frame. Participants were given some time to familiarise themselves with the security challenge, the actors and the technologies involved, the key needs and the main risks associated to it.

Step 2: the group was asked to look into the future and try to imagine how the scenario might have developed in 20 years, considering the needs that could arise for both LEAs and citizens.

Step 3: the group identified which types of changes would have been needed to address the new security needs, considering the technological, organisational and regulatory levels.

At the end, the groups reconvened in the plenary and presented the results of their work.

Participants:

popAI consortium members, Stakeholder Advisory Board, and other external experts.

Outcome

As a final outcome, the workshop produced a set of updated scenarios that will cover a longer timeframe (20 years) and gather initial inputs for the development of the roadmap (e.g. actors, actions, etc.). This information was used in order to set a strategy that would allow to meet the need of the different security scenarios. The strategy indicates:

- the actors involved and their specific responsibility/role;
- the relevant technologies;
- the necessary steps to undertake to ensure such technologies can be used – in a lawful manner – to meet the security needs of each scenario;
- the strategy could indicate – for instance – if specific changes in the legislation are needed, if some actors would have to implement precise actions, etc.

9.1.2 Results of the workshop

9.1.2.1 Scenario 1 "Past will always define future"

Step 1: Presentation of the scenario & Discussion on key points

Participants took a few minutes to familiarise themselves with the scenario and agreed it was quite a complex one, in terms of dynamics and potential issues raised by the actions. In terms of needs for LEAs, they pointed out four main aspects, namely data, citizens' awareness, organisational procedures and ethics.

With regard to data, the LEA representatives highlighted the need for more data, especially if they are to be used for crime prediction. Moreover, it should be recommendable to have more video footage.

They also stress the importance of the use of data: it is crucial to define the purpose of the use of data. The principle of proportionality – as the foundation of any police action – should be strongly embraced. Related to these aspects is awareness: citizens need to know what technologies are used and for which purpose. This would facilitate the acceptance of the technologies and avoid the perception of a police state. As an illustration, if LEAs deploy drones on a neighbour without informing them on the reason, citizens may perceive it as a violation of privacy and react accordingly. If, by contrast, they inform citizens that drones are temporarily deployed to search for a missing kid, they'll much more likely accept the adoption of the technology and this partial violation of privacy.

From an organisational perspective, it would be relevant to map the risks associated with the use of AI-based technologies, as well as create high-level profile to use AI.

The last aspect raised by the group concerned ethics: it is reckoned essential to find a balance between security and privacy. Moreover, in order to be useful and appropriate for the issues potentially raised by AI, ethics should be more robust.

In terms of risks, the group identified violation of privacy as the major risk with, consequently, citizens' mistrust towards the police and the increasing perception of a police state.

In addition, the risk of lack of accountability is quite concerning.

Step 2: Discussion on the scenario in 20 years

The group discussed the potential evolution of such a complex scenario. They agreed on two major developments: law changes and technological innovation.

With regard to the law, regulations would become more robust and the establishment of overseeing bodies is envisaged. For instance, regulations could support the creation of a committee of AI formed by people from variegated backgrounds and areas of society.

From a technological perspective, devices will become more reliable and accurate, thus reducing the number of false positives. In addition, they could be used to support LEAs in raising awareness among citizens. As an illustration, in a 20-year frame, more accurate technologies could be adopted by LEAs for crises, emergencies or investigations.

However, concerns remain regarding the misuse of AI: therefore, ethics should be still considered as the main frame within which to develop AI solutions.

In terms of new needs for LEAs, the group identified the need for more accuracy and immediate response of AI systems, especially if used during emergencies but also the need for legislation to keep pace with the technological development.

However, the human-in-the-loop is still considered essential. AI without human oversight can pose issue of accountability. Human control and oversight are recommendable, as the establishment of ad hoc committees to monitor the use of AI technologies.

From an operational perspective, LEA representatives mentioned that the more frequent and transversal use of AI technologies will ensure less police patrolling; in parallel, training should be strengthened in order to reduce the gap between technology and operator and avoid any kind of human error and/or bias.

Step 3. Identification of future needs (technological, organisational, regulatory)

From a technological perspective, the group identified the main changes needed in relation to the accuracy of AI system and the number of devices used by LEAs. Reliability of technologies is crucial.

With regard to the organizational level, the group stressed the importance of training in order to reduce the possibility of human error; while considering human control and oversight essential.

In terms of regulations, the law must proceed along with technological development to avoid any gaps in the legal framework that could offer the opportunity for misleading use of AI systems. Moreover, provisions for conflict resolution should be included as well.

9.1.2.2 Scenario 2 "AI investigator. Case closed"

Step 1. Presentation of the scenario & Discussion on key points

Participants started analysing the current needs identified in the given scenarios. Such needs can be gathered into three main categories: (a) technological needs; (b) organisational needs; and (c) data-related needs.

In the first category, experts emphasised the need of having equipped and advanced AI systems, such as body-worn cameras. Moreover, they emphasised the importance of digitalising evidence as well as tracking locations.

Concerning organisational needs, AI system training and learning are considered of utmost importance, as well as adopting and providing a clear methodology for it. In addition, the scanning of digital archives could support and facilitate LEAs duties.

Looking at data, the main needs concern the processing, storage and collection of data, thus including issues on the information to collect and the modality of the collection and the type of data.

The discussion, then, proceeded to core risks and the group identified risks related to data, as gaps, partial information and potential errors and biases. Other risks are associated with processes. As an illustration, AI cannot know if somebody changed something (the position of something or data missing for example): hence, the role of the human in the loop is still critical. Secondly, methods for file and evidence sharing should be harmonised and security enforced.

Step 2. Discussion on the scenario in 20 years

The group discussed about the possible development of this scenario in 20 years, while trying to foresee the potential needs that LEAs and citizens may have.

The brainstorming depicted a scenario where the event will be detected by an advanced AI technology that could be a camera or a robot that sends the signals of emergency first. The access and the analysis of the evidence can be done by an AI machine too as well as the processing and storage of data. A sort of private "ChatGPT" could be employed, i.e. specialized AI tools connected to the police that will allow users to establish direct and quick contact with the police – another example could be ALEXA in our homes as a security system to which you can say "alexa, call the police".

Looking from a LEA perspective, the group identified the need of using AI tools in their daily work: however, this should be strongly supported by a specific educational programme as well as a revision and auditing mechanism of the AI tools. Moreover, greater attention should be paid to the social and psychological impacts of these technologies, thus envisaging the introduction of additional qualifications.

Considering citizens, the group suggested the significant advantage of having technologies like Alexa at home connected to the police. This would increase the feeling of security as people would be

directly connected with a police station, thus knowing that someone would be able to promptly react in case of emergency.

At the same time, it becomes necessary to increase trust and transparency in these technologies and their use by law enforcement agencies. In this light, it would be required to conduct and publish auditing as well as provide precise and clear information regarding limitation and regulation of data collection.

Step 3. Identification of future needs (technological, organisational, regulatory)

The group identified four categories of changes necessary in the future: regulatory, technological, organisational, and ethical changes.

In the first category, they include the definition of high-level guidelines, generated in a more public way (e.g. via a more public consultation and a democratic decision-making process), as well as high-level security standards.

In terms of technology, they suggested the need of using visual analytics and adopting an automated auditing system/methodology. Moreover, an impact assessment supported by AI should be available, while the analysis of lessons learnt should be conducted to understand correlations and avoid or reduce main known risks and challenges.

From an organisational perspective, internal and local protocols should be established to ensure that standards and key principles are respected when employing AI-based technologies for security purposes. At the same time, responsible people should have adequate expertise and additional qualifications.

Considering the ethical level, stronger synergies among the multiple stakeholders are necessary, Moreover, in decisions where there are several and possibly divergent interests at stake, representatives from the different categories should be involved in decisions, ranging from the policy level, to LEAs, to citizens.

9.1.2.3 Scenario 3 "Don't shoot the artist"

Step 1. Presentation of the scenario & Discussion on key points

Familiarising with the scenario "Don't shoot the artist" exploring issues regarding cyber operations participants including the ones with technical background, agreed that the scenario looks feasible in general terms. The LEA participants argued on the need for wide and automated access to the raw material of any platform and from any geographical location with a universal authorization/ classification framework. They explained that there are two phases described in the scenario namely, the pre-investigation and after-investigation. Before the investigation of the case, for instance, Open-Source Intelligence (OSINT) can be used. However, after the investigation starts, the forensics procedure needs to be used, otherwise the evidence will be invalid i.e., it cannot be used in a court of law. For example, the entire web representation must be fetched, rendered, and digitally signed to be used in a court of law. Furthermore, the "human-in-the-loop" was emphasized. It was agreed

among the participants that the final assessment of the "red flag" needs to be made by the police officer. LEAs highlighted the need for an explainability framework and for an overall methodology of investigation.

Examining closer the technologies discussed in the scenario, a few points were raised:

- The algorithms used by the police in the scenario are based on algorithms that are approved from a "higher command" entity (e.g., EUROPOL).
- Activity on Dark Web is a priori suspicious.

The points most difficult to be realized are:

- Collecting data from Dark Web unless undercover.
- "Explainable AI" to effectively assess the results of classification algorithms.

Step 2. Discussion on the scenario in 20 years

The discussion regarding the development of the scenario in 20 years' time focused on the fast-pace technological progress. It was claimed that the information/data points will be ubiquitous, in the sense that the focus in the presented scenario was data on social media and dark web, while in 20 years' time data will be produced and collected by smart devices (i.e., IoT, smart homes, smart driving, etc.) and even chips in our bodies. Furthermore, the asymmetry of tech-savviness in the context of LEAs as well as criminals was intensively discussed. For example, it will be easier to identify and arrest criminals who do not have technological capacities than criminal networks that used advanced technologies.

Some critical points were raised. Firstly, the political and border status: the participants expressed concern about the collection and analysis of data outside EU as well as the collection of data in real-time. Secondly, they discussed the modality of tapping these new streams of information. Lastly, the need for LEAs continuous training was emphasized.

Step 3. Identification of future needs (technological, organisational, regulatory)

The discussion about future needs led to three key considerations. The main focus was on LEAs and citizens' training to ensure the correct and lawful use of technologies while increasing awareness across the society. Secondly, the need to avoid digital classification and thirdly the need to increase awareness about the use of generated data, but also the potential risk of malicious use of any public information.

The group discussion concluded that at the present moment, we are far from an explainable AI. The description of gaps from a technological perspective is, thus, of utmost importance: the analysis of big data must be an important issue for LEAs in the near future but it is also crucial to ensure adequate training and knowledge of law data.

9.1.2.4 Scenario 4 "Crossing the invisible borders"

Step 1. Presentation of the scenario & Discussion on key points

The participant of this group first identified some key needs, considering both the perspective of the mission to be accomplished, i.e. border checks and control, as well the travellers' one. They stressed the needs to perform more efficient checks at border control points, being able to handle a large number of travellers. On the other hand, LEAs conducting such checks should ensure less hassle during border control crossing as well as guarantee the safety feeling to travellers and citizens). In addition, a proactive and predictive approach against terrorism should be promoted.

The analysis of the risks pointed out mostly risks related to the travellers and their experience/perception during border control crossing. First, a feeling of discrimination might be perceived based on the type of control LEAs conduct. Secondly, the perception of security could worsen as well as the trust towards the control system.

The possibility of malicious use of data and a lack of transparency may raise ethical concerns and doubts regarding the respect of key principles and rights, such as the right of revocation and oblivion. This may further increase distrust towards the system and those implementing it. Fake news and wrongful accusation may derive from a misuse of data.

Step 2. Discussion on the scenario in 20 years

The group investigated the possible evolution of the scenario in 20 years, depicting an environment where no documents will be needed and the border will be "fully invisible". Passengers will require a pre-approval in order to travel. Passengers without the approval would be denied travelling. In this context, the border crossing will be easier, faster and more open, thus having also a positive economic impact.

Looking into new needs for LEAs, the participants identified the need for more legal and regulatory framework and a better interaction between the AI-technology and the human operator. This could be achieved through improved digital literacy and training. In parallel, the interoperability of systems and infrastructures should be ensured and international collaboration fostered.

At the same time, digital literacy is reckoned essential also for citizens: in this way, their pre-travel experience will be smoother and they will be better aware of processes and the purpose of specific actions/regulations.

Step 3. Identification of future needs (technological, organisational, regulatory)

The new scenario will require some changes in order to ensure LEAs will be able to conduct their tasks.

From a technological perspective, a shift towards more cyber and fewer physical infrastructures is envisaged. A full digital passport will also require an increase in processing power.

In terms of organisational changes, the establishment of cyber risk departments is envisaged in those that previously were only physical departments. Network infrastructures will be improved and new positions will be introduced, such as cyber-crime, digital forensics and digital transformation experts.

Looking into the regulatory level, an update in the GDPR might be required in order to address the new challenges and potential risks related to AI. The approval of the EU AI Act will be essential to have a coherent and common framework that could support the employment of these technologies in security without compromising human rights and fundamental principles.

9.1.2.5 Scenario 5 "Guilty till proven innocent"

Step 1. Presentation of the scenario & Discussion on key points

The key needs emerged from the analysis of this scenario are related to both technological and procedural aspects. Looking at the technological side, accountability and accuracy of the crime data record should be ensured. Algorithms must be developed in a way to avoid discrimination and bias, and takes into consideration the specific context (context-based tools). In addition, transparency is crucial: AI needs to show how the decision has been made.

In terms of procedures, human analysis/support is considered a mandatory requirement. The human operators must be involved in the critical decision-making process, to decide what needs to be further investigated to take an informed decision on the sentencing. Indeed, it is commonly agreed that authorising someone's arrest only based on the data record of an AI tool would not be fair.

The main risks of this scenario are associated with biases, such as direct and indirect discrimination based on biased assumptions (e.g. family ties with someone who has a criminal record). Other risks regard the lack of transparency in an AI-based decision-making process, especially if the algorithm is not context-based. Supervision mechanisms should, then, always be put in place to revise the decision made by the AI tool, to avoid discrimination and/or unfair sentencing; moreover, the human operator should always be kept in the loop.

Step 2. Discussion on the scenario in 20 years

The group, then, proceeded with the foresight activities, drawing the evolution of the scenario in 20 years. In the future environment, algorithms will be more transparent, fair and regulated. In the design phase, experts from ethical legal, and privacy sectors should be involved, as of the early stage of the tool development.

Moreover, a diversification of the technologies based on their use and areas of application might be required. AI technologies that support the investigation of crimes should be considered separately from the technologies used for the prosecution decision-making process. Similarly, technologies that deal with issues of post-sentencing should be kept separated. Despite technological development, however, the intervention of the human judge is still considered critical. In 20 years, there should be a system that takes the best of AI and uses it in the best way: a trained competent human to supervise is still envisaged.

This tool could be helpful to reduce the time and amount of cases that go to Court; however, it cannot replace judges and lawyers who have a lot of arguments to present and who can interpret the law.

In this context, new needs emerged for LEAs include the training of LEAs on a regular basis but also a control mechanism that ensures it is suitable for police officers to use the technology. Independent

assessment must be made upfront. External bodies should provide approval to use a new technology, not to rely fully on the manufacturer

Systems should not be bought without a thorough assessment of the impact and risk. Strong guarantees that technologies will work properly should be given upfront.

Considering the citizens' perspective, needs are mostly associated with respect of fundamental rights and principles. Firstly, citizens must be aware of their rights; in this scenario, Nadia was arrested without knowing what her rights were. Education should provide citizens with enough information on their rights.

There should be a right to challenge the outcomes of AI tools and make a case to review a decision. The fundamental right to access the data that has been fed and processed by the AI tool to enable to make this decision should be ensured. In addition, citizen must know how long their data is being kept, as data retention must be limited in time. Citizens should be made aware of the purpose of the data collection and what it can potentially be used for.

Step 3. Identification of future needs (technological, organisational, regulatory)

The depicted scenario allowed participants to define some required changes.

In terms of technological changes, the group recognised the difficulty in foreseeing any new supporting technology that hasn't been developed yet. The group discussed the possibility of having a tool combining technology and human thinking. The tool would add other possibilities that result from human thinking (e.g. humans could show other potential decisions to the tool; humans, however, should train the tool). The tool would no longer be mono-directional. There would be a counter-AI technology able to contradict the decision of the tool. Such a tool could be trained from real trials or from trial recordings that would be transcripted in text, so arguments are incorporated. The tool could learn from the notes that are taken from the trial.

From an organisational perspective, fundamental changes could be required. As an illustration, additional specialists would be needed in order to train LEAs.

Risks of fake images and fake audio are still perceived as a threat: LEAs will see that crimes are performed with more sophisticated tools. To face this challenge, they will need to understand what type of AI can help conduct policing activities to deal with that.

Decisions taken by AI systems should be reviewed by independent experts, authorized to overturn AI's decision, if necessary. It would require hiring new profiles of people with specific expertise. Police forces should be further trained (about the system development, use of data etc…).

Acquiring new AI capabilities requires also a change in terms of procurement.

Regulatory changes are envisaged as well. The current EU data protection framework, for instance, is blocking the development of AI systems that police officers and LEAs can use because they are not involved as from the development stage.

New regulatory framework should be established to balance privacy, data protection and the need for LEAs to be trained and participate in this development of AI technologies.

An external body would make it a lot easier to train LEAs while complying with the AI act.

Regulations should come first, faster, and be adapted before such technologies are developed. Moreover, it would be desirable to integrate a vision and strategy on the use of the technologies before they are developed.

The group concluded by recognising an increase in the number of experts and judges involved in the process. AI sophisticated tools will be widely used: however, the human component will remain a fundamental part of the process. New technologies will need to be authorised by independent experts. LEAs will be adequately trained and citizens will need to be better aware of threats and also benefits of the technologies.

9.2 The popAI 2040 roadmap

The popAI 2040 roadmap aims to delineate and suggest the actions needed to address the potential scenarios and security risks that LEAs might face in the next 20 years, by relying on an ethical and lawful use of AI-based technologies.

This roadmap takes inputs from the desk research conducted within section3 as well as the inputs from WP3 (Foresight scenarios) and WP4 (Recommendations).

The workshop held in Athens represents a preliminary validation of the roadmap, as the stakeholders agreed on the future scenarios to be taken into consideration when considering the evolution of the European security landscape.

Prior to the illustration of the roadmap itself, it could be relevant to highlight some key considerations that emerged from this analytical work, regarding the technologies, the core types of risks and the identified missing capabilities (Figure 7).

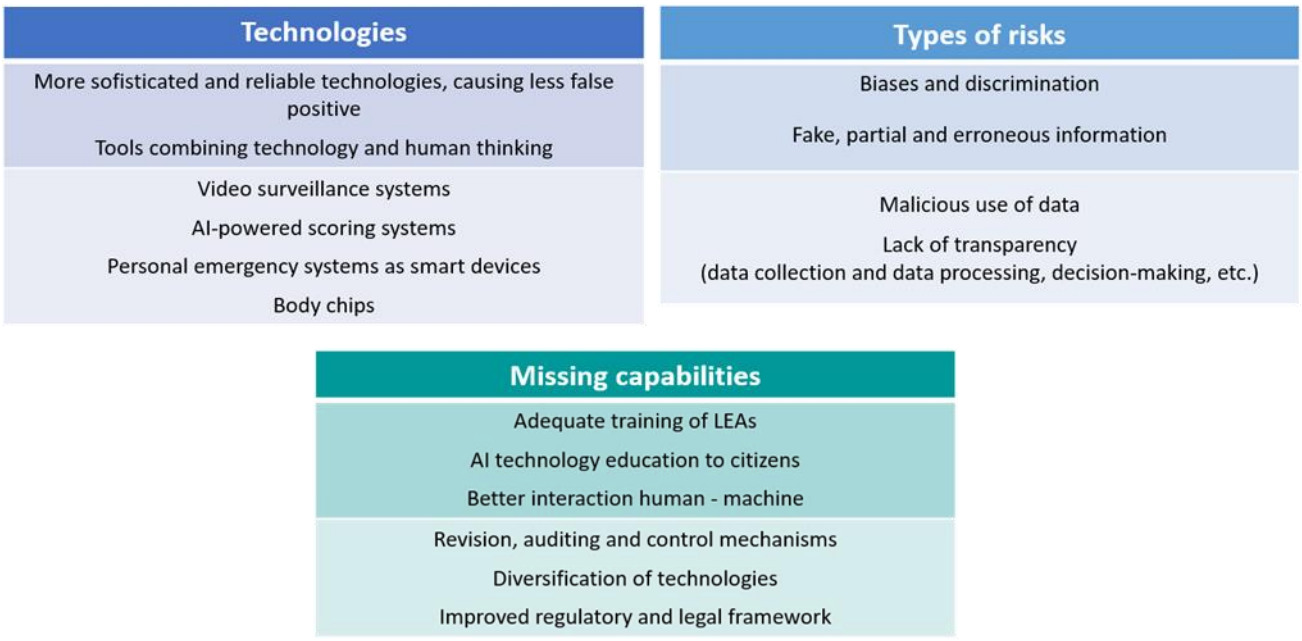

Figure 7 - Key output of the foresight scenarios workshop (Athens)

Based on these considerations, the roadmap includes three strands: a technological strand; an organisational strand; and a regulatory strand. The roadmap is illustrated in Figure 8, while each strand is described more in detail in Table 5.

The technological strand indicates actions that regards the development and/or testing of the future technologies. The organisational strand suggests adjustments that should facilitate the introduction of AI systems into operational procedures, while providing common frameworks and actions. The regulatory strand indicates relevant actions to undertake in order to ensure that the legislative framework keeps pace with the evolution of the technologies as well as the emergence of new societal needs.

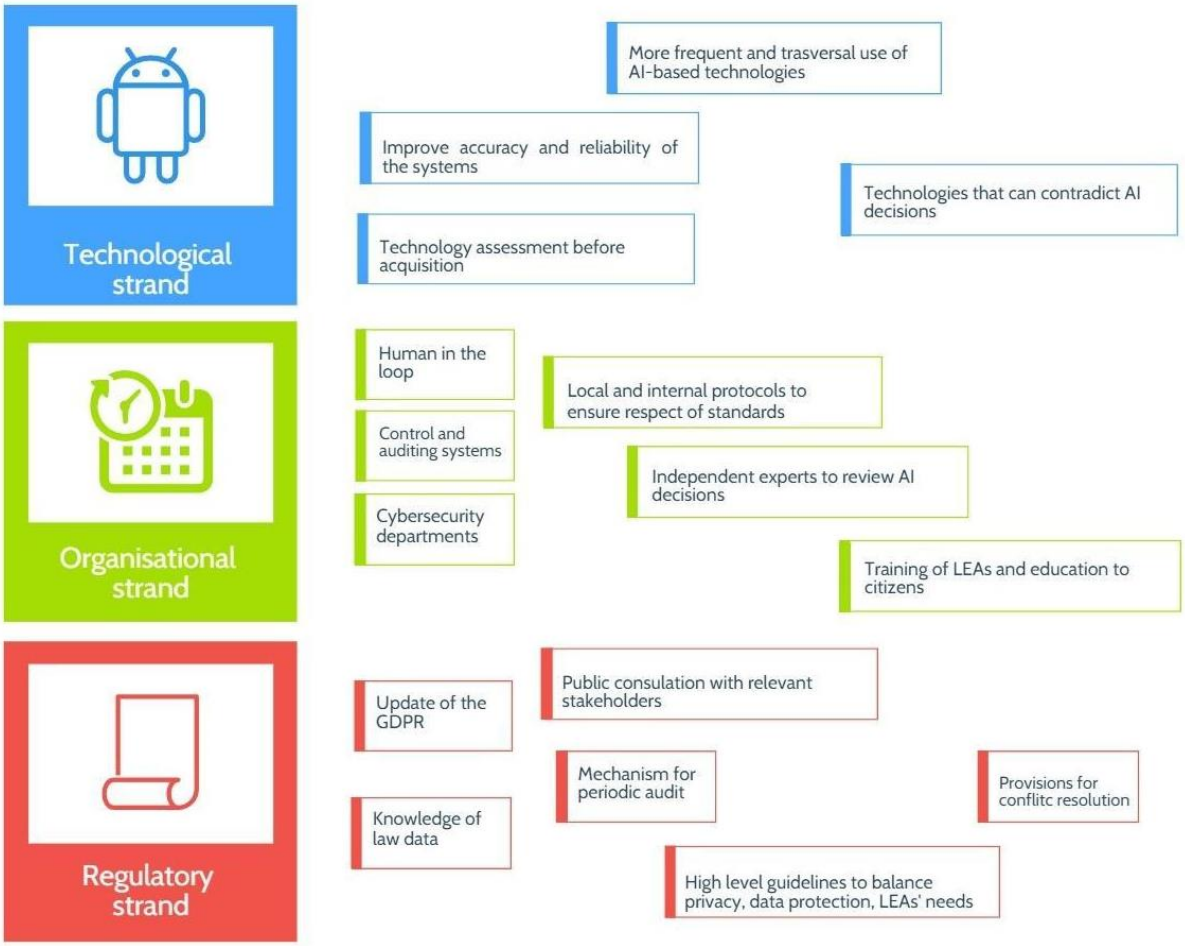

Figure 8 - popAI Roadmap to 2040

Table 5 - popAI Roadmap to (steps)

Technological strand		
Step 1	**Description of the action**	**Targeted stakeholder**

	Improve accuracy and reliability of the system As the reliance on AI-based technology increases, it is fundamental to enhance the accuracy and the reliability of the technologies, thus allowing a lower number of false positives. Common criteria to assess the accuracy of the systems should be established and shared by production companies.	Technology providers; industry
	Assessment of technology before the acquisition Technology providers should undertake tests and provide to end-users an assessment of their system prior to the acquisition. Such assessment should cover the technical performance but also the economic aspect as well as the ethical and environmental impact. Standardised criteria should be adopted throughout the industry to facilitate the comparison and help users to make an informed decision.	Technology providers; industry; users
Step 2	More frequent and transversal use of AI-based technologies The use of AI-based technologies should be encouraged in the different societal sectors, thus contributing to enhance citizens' knowledge, understanding and confidence in these technologies. AI-based technologies adopted across multiple sectors include the analysis of big data and the use of visual analytics. In addition, more cyber infrastructures should be developed.	National authorities, local authorities, critical infrastructure operators (e.g. port operators)
Step 3	Technologies that can contradict an AI decision As the reliance on AI-based technology will increase also in decision-making processes, it is critical to invest also in technologies able to contradict a potential erroneous or vitiated judgement or decision, especially if the results of such AI-based decisions concern sensitive issues.	Technology providers; industry

Organisational strand		
Step 1	Description of the action	Targeted stakeholder
	Human in the loop Human oversight should always be guaranteed even if AI technologies become more reliable and accurate. It is essential, indeed, that the human operator has the final check on the technology as well as the decision based on data or results deriving from the technology.	Technology providers, industry, operators; regulatory authorities

	Establishment of control and auditing systems The employment and use of AI-based technologies should be constantly monitored and controlled. A common mechanism for auditing of technologies should be established, including a common protocol and periodic controls. The outcome of the audit should be examined by relevant regulatory authorities that should get in contact with technology providers and users if something abnormal is detected.	Regulatory authorities; industry; technology providers
	Establishment of cybersecurity department in each organisation Considering the increasing number of cyber infrastructure as well as technologies used within an organisation, cybersecurity departments should be established in each organisation, thus taking care of monitoring and assessing any potential threat as well as ensuring that criteria and standards are constantly met.	Public organisations; private companies; industry; local and national authorities
Step 2	Establish internal and local protocols to ensure respect of standards Organisations should define some internal protocols to guide the implementation of the AI standards when developing and using the technologies. These protocols should ensure full adherence to national and European guidelines in order to avoid conflicting issues or lack of compliance.	Private companies and public organisations; LEAs
	Independent experts to review AI-decisions In line with the provision of always keeping the human in the loop, a group of independent experts should be established with the purpose of monitoring and reviewing the results and decision of AI technologies. Indeed, especially those AI-based systems that are used in decision-making processes are to be carefully monitored to ensure proper procedures are followed, standards implemented and human rights and ethical principles respected.	Regulatory authorities
Step 3	Provide training for LEAs LEAs should be adequately trained not only in the technical functioning of the AI systems but more broadly on the risks it could pose and the measures to overcome/mitigate such risks. Moreover, they should be aware of the legislative framework and the limits of their employment (e.g. banned technologies or restrictions to their use).	LEAs; technology providers; regulatory authorities

	Ensure AI education to citizens Citizens should be trained on AI-based technologies, considering the practical use but also the regulations that must be respected. Moreover, they should be aware of risks deriving from the use of such technologies and informed about any possible applications of these technologies in public areas or for security purposes. This would ensure a more appropriate understanding of the use of AI, thus presumably encouraging the acceptance by citizens.	Citizens; local and national authorities; LEAs

Regulatory strand		
Step 1	Description of the action	Targeted stakeholder
	Update of the GDPR As the evolution of AI technologies continue, it is important to monitor also the provisions of the GDPR. Indeed, some adjustments might be required in the future in order to meet new challenges and risks posed by the AI systems and their possible application in new sectors. Close collaboration with technology providers is highly recommended to facilitate the process and understanding of the needed modifications.	Regulatory authorities; technology providers
	Knowledge of the law It is important also to enhance the knowledge of the law across the society. Increasing awareness among citizens regarding rules but also rights should help to ensure that the use of AI technologies in all the societal sectors occurs in compliance with the existing regulatory framework, and does not contravene fundamental rights.	Citizens; regulatory authorities; local and national authorities
Step 2	Public consultation with relevant stakeholders Technologies evolve rapidly and law usually struggles to keep pace. It is crucial to have periodic public consultations with relevant stakeholders to understand challenges and risks related to the use of AI-based technologies. This approach would allow a greater understanding of the landscape as it evolves, thus permitting regulatory authorities to proactively adjust the legislative framework to meet new challenges and needs.	Regulatory authorities; local and national authorities; civil society; technology providers; industry

	Mechanism for periodic audit Common procedures for periodic audit should be established. This would meet a two-fold goal: on the one hand, it will allow to monitor that rules and standards are respected; on the other, it would offer an opportunity to identify challenges and risks, thus having the chance to prevent the emergence of irregularities.	Regulatory authorities
Step 3	Definition of high-level guidelines to balance privacy, data protection, and LEAs' needs The evolution of AI systems and their use by LEAs for security purposes might create challenges in terms of privacy and data protection. The public consultations should allow to understand also potential clashes between the needs and priorities of the different stakeholders. Stemming from this, high level guidelines should be defined to balance LEAs' needs in the use and application of AI-based technologies for their activities and privacy and data protection for citizens.	LEAs; citizens; regulatory authorities
Step 4	Provisions for conflict resolution The evolution of the regulatory framework should also include provisions for conflict resolution, in case of breach of any rule or limitations. Such provisions should be implemented by a dedicated entity in charge of assessing the conflict and deliberate the course of action.	Regulatory authorities; local and national authorities

10 Conclusions

This report describes the outcome of Task 5.5 dedicated to the definition of two roadmaps, a short-term roadmap and a long-term one.

Part A addresses the short-term compliance and certification roadmap that aims pave the path to a "European common approach" for compliance and certification of AI-based technologies used in the frame of Law Enforcement activities.

To achieve this goal, a preliminary analysis of the existing standards, certifications, guidelines and overall regulatory frameworks has been performed, thus delivering a "catalogue" where all these documents are organised according to three levels (national, European and international) and mapped into the ALTAI principles.

The roadmap outlines six steps: 1. coordination at national and European levels; 2. harmonisation of language and frame of actions; 3. identification of gaps; 4. guidelines to fill the gaps; 5. training for LEAs and education to citizens; and 6. monitoring and update. Each step is described and the target audience indicated.

Part B addresses the long-term roadmap to 2040, which highlights potential futures scenarios and risks, trying to understand the role that AI-based technologies might play in these futures and identifying strategies to get there.

To define the roadmap, a preliminary analysis of the current EU security landscape and the strategic priorities of the European Union was conducted. In addition, the roadmap receives the inputs from WP3 (foresight scenarios) and WP4 (recommendations).

The final roadmap combines three strands: a. the technological strand, indicating actions that regards the development and/or testing of the future technologies; b. the organisational strand, suggesting adjustments that should facilitate the introduction of AI systems into operational procedures, while providing common frameworks and actions; and c. the regulatory strand, indicating relevant actions to undertake in order to ensure that the legislative framework keeps pace with the evolution of the technologies as well as the emergence of new societal needs.

Each strand envisages multiple steps to be carried out either in parallel or in a consequential fashion. Similar to the short-term roadmap, the 2040 roadmap indicates also the relevant target audience.

11 References

[1] European Parliament, «EU AI Act: first regulation on artificial intelligence,» European Parliament News, 14 06 2023. [Online]. Available: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence . [Last accessed on 22 08 2023].

[2] High-Level Expert Group on Artificial Intelligence, «Ethics Guidelines for Trustworthy AI,» European Commission, Brussels, 2019.

[3] High-Level Experts Group on Artificial Intelligence, «The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment,» European Commission, Brussels, 2020.

[4] United Kingdom Government, «Gov.uk Service Manual,» 16 11 2017. [Online]. Available: https://www.gov.uk/service-manual/agile-delivery/developing-a-roadmap. [Last accessed on 21 08 2023].

[5] European Commission, «European Standards,» European Parliament Internal Market, Industry, Entrepreneurship and SMEs, 22 08 2023. [Online]. Available: https://single-market-economy.ec.europa.eu/single-market/european-standards_en. [Last accessed on 22 08 2023].

[6] A. B. e. a. Cremers, «Trustworthy use of Artificial Intelligence. Priorities from a philosophical, ethical, legal, and technological viewpoint as a basis for certification of artificial intelligence,» Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin (Germany), 2019.

[7] AI TWG (Transnational Working Group), «Report of TWG AI: Landscape of AI Standards,» 2023.

[8] W. Ziegler, «A Landscape Analysis of Standardisation in the Field of Artificial Intelligence,» *Journal of ICT,* vol. 8, n. 2, pp. 151-184, 2020.

[9] European Commission - Internal Market, Industry, Entrepreneurship and SMEs, «European Standards,» European Commission, [Online]. Available: https://single-market-economy.ec.europa.eu/single-market/european-standards_en. [Last accessed on 23 08 2023].

[10] AI Ethics Impact Group, «From Principles to Practice: An interdisciplinary framework to operationalise AI ethics,» Bertelsmann Stiftung, Gütersloh, 2020.

[11] L. T. Wiehler, *How can AI Regulation be Effectively Enforced? Comparing Compliance Mechanisms for AI Regulation with a Multiple-Criteria Decision Analysis,* European University Institute - School of Transnational Governance, 2022.

[12] M. e. a. Poretschkin, «Guideline for Designing Trustworthy Artificial Intelligence,» Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, 2023.

[13] European Foresight Platform (efp), «Scenario Method,» European Foresight Platform, [Online]. Available: http://foresight-platform.eu/community/forlearn/how-to-do-foresight/methods/scenario/ . [Last accessed on 24 08 2023].

[14] European Commission, «Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy,» European Commission, Brussels, 2020.

[15] H. a. A. W. Kahn, The Year 2000: A Framework for Speculation on the Next Thirty-three Years, New York: Macmillan, 1967.

[16] L. Freedman, Strategy, New York: Oxford University Press, 2013.

[17] World Economic Forum, «Global Risks Report 2023,» WEF, Davos, 2023.

[18] European Commission, «EU Security Union Strategy,» Brussels, 2020.

[19] European Commission, «A European approach to artificial intelligence», European Commission, [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence [Last accessed on 22 09 2023].

[20] European Commission, «2023 State of the Union Address by President von der Leyen», European Commission, [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/speech_23_4426 [Lasst accessed on 22 09 2023].

Annex I (Part A)

This annex includes additional information on the concept of standard (see section 1.2).

- **Standard**

The standardisation process unfolds through the collaboration of technical commissions, market regulations, stakeholders' cooperation, insights from experts in diverse fields, and consensus-building approach. Standards, in this context, are voluntary and consensus-based, offering specifications and test methods.

It should be mentioned that standards are formulated and developed by independent organisations, which can be categorized as either Standard Setting Organisations (SSO) or Standard Development Organisations (SDO).

Standard Setting Organisations (SSO) are those entities "*whose primary activities are developing, coordinating, promulgating, revising, amending, reissuing, interpreting, or otherwise maintaining specifications and standards that address the interests of a wide base of users outside the standards development organization"*. Examples are the World Wide Web Consortium (W3C) and the Internet Research Task Force (IRTF).

Standard Development Organisations (SDO), instead, are *Standard Setting Organizations that have a formal recognition by international treaties, regulation, etc.* Hence, SDOs are a subset of SSOs. The International Electrotechnical Commission (IEC), the International Organisation for standardisation (ISO) and the International Telecommunication Union (ITU) are international SDOs [8].

The European Standard Organisation (ESO) is acting as a European platform through which European Standards are developed. The three ESOs are the European Committee for Standardisation (CEN) and the European Electrotechnical Committee for Standardization (CENELEC), the European Telecommunications Standards Institute (ETSI). Only standards developed by the three ESOs are recognized as European Standards (ENs) [9].[3]

On a final note, it is worth clarifying that *Standards* are the outputs from an SDO, while *Specifications* are outputs from an SSO that may become standards when ratified by an SDO [8].

Annex II (Part A)

SSO (Standard Setting Organisation): any entity whose primary activities are developing, coordinating, promulgating, revising, amending, reissuing, interpreting, or otherwise maintaining specifications and standards that address the interests of a wide base of users outside the standard development organisation.
W3C: While W3C has a number of published standards and such under development that are not explicitly developed for AI but are well suited for usage in the AI field (e.g. data description or formatting standards). The group studies AI knowledge representation. No standards development is planned. **IRTF:** it is composed by three research groups working on studies on the possible role of AI in future networks. No standards development is planned

[3] Annex I provides more information on these organisations.

SDO (Standard Development Organisation)- (International): a standard setting organisation that has a formal recognition by international treaties, regulation etc. SDOs are a subset of SSOs.
ISO: International Standards Organization with a membership of 167 national standards bodies. Members are the foremost standards organisations in their countries and there is only 1 member per country (Members Portal: https://www.iso.org/members.html). It does not perform certification and does not issue certificates. Certification is performed by external certification bodies, thus companies or organizations cannot be certified by ISO. The ISO's Committee on Conformity Assessment (CASCO) has produced several standards related to the certification process, which are used by certification bodies (CASCO standards) **IEC:** International Electrotechnical Commission is a global, not-for-profit membership organization, whose work underpins quality infrastructure and international trade in electrical and electronic goods. The IEC brings together more than 170 countries and provides a global, neutral and independent standardization platform to 20 000 experts globally. It administers 4 Conformity assessment systems whose members certify that devices, systems, installations, services and people work as required. **IEEE:** IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE SA (IEEE Standardiasation Associations) nurtures, develops, and advances the building of global technologies. As a leading developer of industry standards in a broad range of technologies, IEEE SA drives the functionality, capabilities, safety, and interoperability of products and services, transforming how people live, work, and communicate. **ITU:** The International Telecommunication Union (ITU) is the United Nations specialized agency for information and communication technologies – ICTs. Founded in 1865 to facilitate international connectivity in communications networks, we allocate global radio spectrum and satellite orbits, develop the technical standards that ensure networks and technologies seamlessly interconnect, and strive to improve access to ICTs to underserved communities worldwide.
ESO (European Standards Organisation)
CEN: European Commitee for Standardization **CENELEC**: European Committee for Electrotechnical Standardization **ETSI**: European Telecommunication Standards Institute.
National normalisation and certification entities
1. Italy **UNI** Ente Nazionale Italiano di Unificazione. Elabora le norme italiane, collabora con gli enti normatori internazionali, concede l'eventuale marchio UNI a prodotti conformi a determinate norme **CEI** Comitato Elettrotecnico italiano, opera in analogia e collaborazione con l'eventuale marchio UNI nel settore elettrico **2. Belgium** **NBN** Bureau de normalisation Bureau voor Normalisatie **3. Bulgaria** **БИС** Български институт за стандартизация **4. Czech Republic** **ÚNMZ** Úřad pro technickou normalizaci, metrologii a státní zkušebnictví **5. Denmark** **DS** Fonden Dansk Standard **6. Germany** **DIN** Deutsches Institut für Normung e.V. **DKE** Deutsche Kommission Elektrotechnik Elektronik Informationstechnik im DIN und VDE **7. Estonia** **EVS** Eesti Standardikeskus **TJA** Tehnilise Järelevalve Amet **8. Ireland** **SAI** National Standards Authority of Ireland **9. Greece** **ΕΣΥΠ/ΕΛΟΤ** Εθνικό Σύστημα Υποδομών Ποιότητας/Αυτοτελής Λειτουργική Μονάδα Τυποποίησης ΕΛΟΤ

10. Spain
AENOR Asociación Española de Normalización y Certificación
11. France
AFNOR Association française de normalisation
12. Croatia
HZN Hrvatski zavod za norme
13. Cyrpus
CYS Κυπριακός Οργανισμός Τυποποίησης (Cyprus Organisation for Standardisation)
14. Latvia
LVS Latvijas standarts
15. Lithuania
LST Lietuvos standartizacijos departamentas
16. Luxembourg
ILNAS Institut luxembourgeois de normalisation, de l'accréditation, de la sécurité et qualité des produits et services
17. Hungary
MSZT Magyar Szabványügyi Testület
18. Malta
MCCAA L-Awtorita' ta' Malta għall-Kompetizzjoni u għall-Affarijiet tal-Konsumatur
19. Netherlands
NEN Stichting Nederlands Normalisatie-instituut NEC Stichting Nederlands Elektrotechnisch Comité
20. AustriaASI Austrian Standards Institute (Österreichisches Normungsinstitut)
OVE Österreichischer Verband für Elektrotechnik
21. Poland
PKN Polski Komitet Normalizacyjny
22. Portugal
IPQ Instituto Português da Qualidade
23. Romania
ASRO Asociația de Standardizare din România
24. Slovenia
SIST Slovenski inštitut za standardizacijo
25. Slovak
SÚTN Slovenský ústav technickej normalizácie
26. Finland
SFS Suomen Standardisoimisliitto SFS ry; Finlands Standardiseringsförbund SFS rf
FICORA Viestintävirasto Kommunikationsverket
SESKO
Suomen Sähköteknillinen Standardisoimisyhdistys SESKO ry; Finlands Elektrotekniska Standardiseringsförening SESKO rf
27. Sweden
SIS Swedish Standards Institute
SEK Svensk Elstandard
ITS Informationstekniska standardiseringen
28. United Kingdom
BSI British Standards Institution
29. USA

ASA: American Standard Association. It represents USA at ISO

Annex III (Part B)

Annex III.1 Scenario 1 "Past will always define future"

Security threat	**Technology**
Crime prevention Predictive policing	AI-powered surveillance systems
Description of the scenario	
AI algorithms for civil security purposes use police data, combined with other datasets such as demographic, abstracted data from mobile phones, and socio-economic data, as well as data that come from hotspot methods to predict when and where criminal activities are most likely to occur. Interoperability of diverse data sources is authorised in support of crime prevention and community safety. Several local 'blacklists' have been created among European Member States that can be linked, compared, and updated in a European level. Based on advanced algorithmic processes, AI-powered surveillance systems are installed in areas flagged as high-risk while drones often circle over. Federico is an Italian political activist. He has studied chemistry but is unemployed. When he was a teenager Federico was a musician and through his music, he was protesting xenophobia and racism. Due to his beliefs, he was often victim of far-right extremists. He never gave up on his ideas. Last year, Federico visited some family friends in Barcelona with his parents for two weeks. During their stay, his mother was feeling rather weak and therefore they mainly relaxed at their friend's hotel without visiting tourist attractions. At the same time of the year, in Spain's capital, there were riots on the streets against austerity. Several people were prosecuted. On their way back to Italy, Federico and his parents were asked a few questions by the airport security staff. Two weeks after their return in Italy, Federico bought online a ticket for a big concert that didn't match his music taste. Political figures from the government would also attend this concert. In the same afternoon, Federico joined a telegram group calling for action against European austerity policies. Some of the concert's technicians were also members of this group as well as left wing extremists. The night of the concert Federico noticed that a drone was following him. He had already a difficult day. Suspecting his past might be still triggering algorithmic systems to surveil him he gets angry. The sensors in his car record Federico's tension. The algorithm flags Federico as a high-risk case. The AI-powered system sends a signal to the next available operational unit based on the distance as well as their available equipment, skills, and experience. A police car approaches him a few minutes later and the police officer asks him to follow them to the nearest police station. Federico reacts but complies with the request. Federico is soon released as his case was a false positive. Police officers insert the new data in the system and Federico's scoring is updated.	

Annex III.2 Scenario 2 "AI investigator. Case closed"

Security threat	**Technology**
Crime investigation	AI-powered ranking system AI-based CCTV camera
Description of the scenario	
AI algorithms for civil security purposes use police data, combined with other datasets such as demographic, abstracted data from mobile phones, and socio-economic data, as well as data that come from hotspot methods to predict when and where criminal activities are most likely to occur. The interoperability of diverse data sources is authorised in support of crime prevention and community safety. Several local 'blacklists' have been created among European Member States that can be linked, compared, and updated in a European level. Based on advanced algorithmic processes, AI-powered surveillance systems are installed in areas flagged as high-risk while drones often circle over. Federico is an Italian political activist. He has studied chemistry but is unemployed. When he was a teenager Federico was a musician and through his music, he was protesting xenophobia and racism. Due to his beliefs, he was often victim of far-right extremists. He never gave up on his ideas. Last year, Federico visited some family friends in Barcelona with his parents for two weeks. During their stay, his mother was feeling rather weak and therefore they mainly relaxed at their friend's hotel without visiting tourist attractions. At the same time of the year, in Spain's capital, there were riots on the streets against austerity. Several people were prosecuted. On their way back to Italy, Federico and his parents were asked a few questions by the airport security staff. Two weeks after their return in Italy, Federico bought online a ticket for a big concert that didn't match his music taste. Political figures from the government would also attend this concert. In the same afternoon, Federico joined a telegram group calling for action against European austerity policies. Some of the concert's technicians were also members of this group as well as left wing extremists. The night of the concert Federico noticed that a drone was following him. He had already a difficult day. Suspecting his past might be still triggering algorithmic systems to surveil him he gets angry. The sensors in his car record Federico's tension. The algorithm flags Federico as a high-risk case. The AI-powered system sends a signal to the next available operational unit based on the distance as well as their available equipment, skills, and experience. A police car approaches him a few minutes later and the police officer asks him to follow them to the nearest police station. Federico reacts but complies with the request. Federico is soon released as his case was a false positive. Police officers insert the new data in the system and Federico's scoring is updated.	

Annex III.3 Scenario 3 "Don't shoot the artist"

Security threat	**Technology**
Cyber Operations	AI system for web crawling

Description of the scenario

Crimes of child pornography and exploitation have been rising with the increased use of the internet and the widespread use of the dark web. At the same time, the cases of human operators experiencing post-traumatic disorder and other mental health issues due to daily exposure to child pornography are rising dramatically. Therefore, LEAs have been using an AI system that crawls the web, including social media sites, for images of child sexual abuse. The system allows automated processing, assessment, and prioritisation of child sexual abuse material (CSAM). In addition, once such material is flagged the system records the 'journey' of the material and identifies all internet users, including dark web and peer-to-peer file sharing networks, who interacted with it, including posting, reposting, downloading, saving, processing and so on.

The system then runs an automatic crawling of online sources for complementary information for investigations in compliance with the national legal requirements and provides a score flagging those representing a high risk. The algorithm that provides the scoring is based on their history, online activity, and other factors such as demographics, network, and others. The criteria used by the algorithm are not public. LEAs have access to private databases for the flagged users.

The use of the system has proved efficient in many cases and now, the human operators have to assess much less volume of child abuse material, especially in cases of objection to the automated results and further investigations. A huge volume of such material has been removed from the internet and many abusers are jailed.

John is a 42-year-old Englishman who has moved to Greece since Brexit. John works as a photographer. He is homosexual and last year he adopted a 2-year-old child with his partner. John mainly promotes his work through social media such as Instagram, TikTok, and YouTube. He shares photos, as well as snapshots "behind the scenes" sharing photography tips. John is inspired by the seaside. This is why he chose to live on a small Greek island.

Since he became a father though, his main inspiration are the children and their relationships with adults, with the environment, and so on. In this context, he shares pictures online depicting young children in swimsuits with adults nearby. Recently, he joined online communities for parents and children. He is preparing an exhibition on the empowerment of children through photography and conducts some research.

The automatic system falsely identifies some of his photos as CSAM as an algorithm embedded in the web crawler proved unfairly biased against specific characteristics – sexual orientation, age, background, etc. All of his photos are removed, and his accounts are suspended. A police officer appears at John's house and takes him to Athens for further investigation. He is falsely accused, and these accusations have terrible effects on his work and life. Even though he is discharged, this whole situation has ruined both his professional fame and his relations on the small island. He and his family decide to move to another place, and he slowly starts working again using a nick name. Along with

other photographers, cartoonists, and other artists, they form a campaign group to make the algorithm fairer.

Annex III.4 Scenario 4 "Crossing the invisible borders"

Security threat	**Technology**
Migration, Asylum and border control	AI-based intelligent video surveillance system

Description of the scenario

Brussels' airport has installed an AI-based intelligent video surveillance system to monitor travellers' entire trip from check-in to boarding, using solely their face as a form of identification. The system uses a facial recognition system with CCTV cameras installed in the airport. Biometric templates created with the camera footage are used for comparison with the travellers' passports. Besides, the system monitors behaviour within the border control areas, with the purpose of producing warnings for potential anomalies and suspicious events. The system also analyses a combination of behaviour and appearance risk indicators which contribute to an aggregated risk calculation, from both negative and positive indicators. In cases where the system is triggered, the biometric templates are also compared to datasets of criminals and suspects of crime.

Joe enters Brussels' airport to catch his return flight home after a business trip. He is a journalist in the Netherlands. In Brussels, he covered a special European Council meeting regarding EU migration and asylum policy. Joe is himself a migrant from Syria. His family managed to migrate when Joe was just two years old. Even though he only briefly lived in Syria, he was often treated differently because of his ethnicity. He managed to study nevertheless and for the last three years, he has been working as a freelance journalist.

Joe arrives early at the airport and instead of proceeding to the security check, he wanders around the arrivals area as he talks on the phone with a colleague. Without realising it and in pursuit of a quiet place, he walks just inside of a restricted area of the airport as he makes phone calls and checks his messages, emails etc.

Video surveillance analysis based on AI triggers and raise an alert based on a combination of risk indicators triggered by his appearance, behaviour, and current location. The alert activates the process of automated analysis across multiple datasets. Joe's full history comes up including his passport information, articles he has published, public posts on his social media, as well as CCTV footage from the demonstrations outside the European Parliament where the spJoe walks towards his gate where he attempts to scan his ticket. However, his attempt fails and, in the meantime, a security officer appears and asks him to follow her/him. Joe is not surprised as he is aware of the AI-based intelligent video surveillance system installed in the airport. Using his journalist hat, he is asking for a report on the algorithmic decision. The officer cannot disclose the AI explainability report he received as the indicators are classified on the basis of public safety. However, he displays the EU certification which has assessed and validated the AI system as operating in a responsible and trustworthy manner. Joe is filing an official report asking for full disclosure before he continues his journey.

Annex III.5Scenario 5 “Guilty till proven innocent”

Security threat	**Technology**
Administration of Justice	AI-powered scoring system
Description of the scenario	
AI systems have been gradually employed in the courts of the European Member States. Indeed, the use of AI to support the decision making at every stage of the criminal justice system is encouraged given the large number of cases to be judged. In this line, algorithmic tools have been assisting the decision-making process on whether a prosecuted person should be immediately released as innocent, if they should have a financial penalty, or the case should be assessed in court. The AI system at this stage is built on data from diverse sources, including the history of the prosecuted persons that exist in police database, the national databases, as well as all the evidence collected throughout the investigation process. Data can be completed by social media and the web depending on the seriousness of the crime. If the case goes to court, the system is further fed with the evidence presented in court in real time. At the end of the hearing process, the system makes the calculations based on all the data, looking for patterns and comparing the case with similar past ones. Finally, it suggests to the judge the risk of reoffending within the following five years and indicates if the risk for an individual is low, medium, or high. The scoring is accompanied by a report that indicates the data and the criteria based on which the score emerged. Nadia is approaching a jewellery store when a man passes by her, falling into her in his rush. She ignores the situation and enters the store to buy a present for her mother’s birthday. As soon as she enters the security door closes behind her and a police officer arrests her. Nadia is totally confused. She tries to protest but everything happens very quickly. In her bag, they find a stolen ring with a diamond. She knows she did not steal the ring, but she cannot prove it. Nadia has been raised in a rather problematic household. Her father was an alcoholic with a history of committing intimate partner violence. They were often in trouble with the police. She knows that she has a police record even though she was the victim. Similarly, as a teenager, she also ended up at the police station following a fight with some girls that were bullying her at school. The AI system assigned her a high-risk scoring suggesting two years in prison. Nadia objected and her lawyer asked for the CCTV footage of the area. The scene where the man falls into her while he is exiting the jewellery store is captured. The system runs a check using the facial recognition to compare the man’s face with other databases. The person is identified but the system gives low scoring. He is a middle-class businessman with no record with the police.	

Annex IV (Part B)

Annex IV.1 Recommendations for LEAs

The recommended practices start with the necessity for law enforcement AI to be designed and developed following an ethics-by-design, privacy-by-design and security-by-design approach. Then, the importance of AI literacy, including education and training of LEAs on the use of AI-enabled technologies, was highlighted and relevant information about the types, frequency and implementation of the educational and training courses was provided. The popAI Ethics Toolbox (developed within WP2) was presented as an example of adequate training of LEAs on AI. Furthermore, considering the various potential risks and impact of law enforcement AI on the affected persons, emphasis was put on the different types of impact assessments that have to be conducted prior to the deployment of AI systems by LEAs. Additionally, to ensure the ethical and transparent use of AI by LEAs, emphasis was put on the affected persons by providing recommendations related to the inclusion of the civil society in the AI process and giving examples of awareness raising and transparency tactics that will enhance the role of citizens and will increase their trust. Finally, the establishment of a relevant AI department or committee that consists of LEAs, ethics and legal experts, policymakers and technology developers (to be in communication with the civil society) was highly recommended to ensure that AI is used in law enforcement in an ethical and lawful way by taking into consideration the knowledge and perspectives of all parties involved in the AI lifecycle.[4]

Annex IV.2 Recommendations for policymakers

The harmonisation of the regulatory framework at the EU level is suggested, considering, in addition, the applicable data protection framework and the adaptation of the legislation to the technological developments, and the adoption of "Recommendations" in complementarity to the AI Act. Institutional safeguarding and procedure building is proposed, and specifically the establishment of training and awareness raising educational programmes under the AI Literacy notion, the promotion of EU cooperation among stakeholders, the institutionalisation of multidisciplinary collaboration, the standardisation of the Impact Assessment procedure and the AI procurement procedure, the establishment of transparency and accountability protocols for LEAs, through a well-designed procedure (e.g. a platform) for the exercise of the citizens' rights, and the investment of EU-funding on the ethical and legal AI research, design, development and deployment.

[4] popAI D4.1 'White Paper for LEAs'

Although AI technologies can provide a valuable support to LEAs in the exercise of their operational functions, a balancing exercise between authorising the use of AI and protecting human rights and freedoms must be conducted. While the training of LEAs is a recurrent trend that appears to have been broadly agreed upon by the different stakeholders involved in this exercise, using AI in an ethical manner that supports the work of LEAs and yet respects privacy cannot be achieved without a solid legislative framework. Legislation at the EU level setting the conditions for and limits concerning the use of AI by LEAs should be adopted in accordance with the rule of law. Ensuring a non-discriminatory use of AI tools can be accomplished through appropriate training of LEAs as mentioned above. In addition, combatting unfair bias shall be part of the AI system's design stage. Further on, from the early stage of conception, AI tools to be used by LEAs should be periodically and systematically assessed (with respect to their social impact and/or impact on rights and fundamental freedoms) to ensure fair and ethical suggestions that -among others- they do not target citizens on the basis of protected grounds of discrimination. Moreover, greater awareness among citizens and participation in the impact assessment stage should be promoted. In a nutshell, a holistic approach is recommended, combining rigorous training of LEAs, robust regulation of AI systems, citizen participation and design and development in a manner that upholds ethical and legal principles and safeguards societal values. [5]

Annex IV.3 Recommendations for technology developers

Recommendations for Technology Developers regarding the ethical use of AI by LEAs were indicatively categorised as: recommendations regarding the stage of design of AI systems, the development of AI systems, the processing of data and horizontal recommendations for technology developers regarding the ethical use of AI for LEAs, which are of a more general nature. [6]

[5] popAI D4.2 'White Paper for the Civil society'

[6] popAI D4.3 'White Paper for Technology Developers'