AP4AI
Accountability Principles for AI

Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain

Summary Report on Expert Consultations

Accountability Principles for Artificial Intelligence (AP4AI)
in the Internal Security Domain
Summary Report on Expert Consultations

Version 27 January 2022

Coordinated by:

- Europol Innovation Lab
- CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research)

Supporting Partners:

- Eurojust
- EUAA
- CEPOL

Disclaimer: This report is intended as a reference document to establish the foundation of Accountability for AI in the Internal Security Domain. It defines the initial set of 12 AP4AI Principles developed by the AP4AI consortium in consultations with international AI experts in the first phase of the project in combination with a review of existing AI frameworks published in the last six years. The report further provides an account of the methodological approach, as well as an outline of upcoming steps in the project. This first Summary Report further serves to demonstrating progress of the project and as vehicle to continue the collection of feedback and insights from experts across pertinent fields. The project received ethics approval by the university ethics board of Sheffield Hallam University, where CENTRIC as academic lead of the AP4AI Project is located. All participants were asked to give their written informed consent before providing written input and take part in the consultation sessions. The informed consent also requested explicit permission for the recording of the consultation sessions, their transcription and for quotes to be used in anonymised form. The AP4AI Project is jointly conducted by CENTRIC and Europol and supported by Eurojust, EUAA and CEPOL with advice and contributions by the EU Agency for Fundamental Rights (FRA), in the framework of the EU Innovation Hub for Internal Security. The research outcomes, the opinions, critical reflections, conclusions and recommendations stated in this report do not necessarily reflect the views of CENTRIC, Europol, FRA, CEPOL, EUAA or Eurojust.

Authors:

- B. Akhgar, P.S. Bayerl, K. Bailey, R. Dennis, S. Heyes, A. Lyle, A. Raven, F. Sampson, *CENTRIC*
- M. Gercke, *Cybercrime Research Institute*

Acknowledgement
The AP4AI Project would like to express its appreciation to the large number of experts who provided their time and insight for this work.

FOREWORD

The European Police Code of Ethics states that, "the police role in upholding and safeguarding the rule of law is so important that the condition of a democracy can often be determined just by examining the conduct of its police."[1] This powerful statement sits at the heart of accountable policing, and the accountability of all organisations for which security and justice is at the core of their mandate.

Accountability is easily claimed, yet hard to evidence, and the use of Artificial Intelligence (AI) brings a particular dilemma. On the one hand, the opportunities offered by burgeoning AI technologies present law enforcement agencies (LEAs) and justice organisations with sweeping new capabilities to make sense of vast amounts of information to quickly find the missing, protect the vulnerable and monitor possible threats. For example, the combination of Natural Language Processing (NLP), which enables computers to analyse and process large volumes of natural language-based data, and computer vision, which allows object detection and image classification, has revolutionised traditional criminal analysis by allowing analysts to quickly extract relevant information, identify relations and make connections that were not humanly possible before. Some AI models provide the potential to solve old crimes by helping investigators to identify cases containing promising evidence that could be re-examined using new forensic techniques. They also help to prevent new crimes, particularly where technology is being used by the perpetrators, for example, to detect cyber-attacks faster, identify large-scale frauds or fend-off disinformation campaigns. AI and ML tools increase the speed, accuracy and range of investigations such as combating child sexual exploitation while at the same time reducing costs and redirecting resources more efficiently and effectively. AI capabilities can also enhance the criminal justice system by providing risk assessment tools based on historical data of cases and judgments, which bring additional pieces of information a judge may take into consideration. AI tools can, for instance, assist judges in assessing an individual's risk of reoffending or the risk of failure to appear in court during the pre-trial phase.

On the other hand, there is evidence of citizens' concerns about the use of AI by LEAs, and the internal security community more generally. Overly relying on historical arrest and crime data as a predictive factor may lead to biased results, perpetuating historical bias in security practices and undermining efforts to ensure individualized and equal justice.[2] This is coupled with a growing propensity to challenge AI use in court utilising a complex network of laws focused on the protection of privacy and other fundamental rights and freedoms.

Clearly, solutions are needed that help to achieve a healthy balance between the need of security practitioners to innovate their practices and use modern technologies to enhance capabilities in order to remain effective against a developing crime landscape on the one hand, and the legitimate expectation by citizens that police work is conducted lawfully, proportionately and in pursuit of a legitimate aim, including also the ability to ask for consequences and remedies if things go wrong.

In 1829 Sir Robert Peel stated that police need "*to recognise always that the power of the police to fulfil their functions and duties is dependent on public approval of their existence, actions and behaviour, and on their ability to secure and maintain public respect.*"[3] The AP4AI Project represents an attempt from a group of European Union Agencies and their research partner CENTRIC, in the framework of the EU Innovation Hub for Internal Security, to support this important aspiration to police and justice accountability, grounded in fundamental EU values. With its ambitious scope, AP4AI will assist the internal security community in achieving sustainable and accountable security innovations with confidence, and most importantly, with confidence from the society it seeks to protect and serve.

Gregory Mounier
EU Innovation Hub Team
Europol

Babak Akhgar
Director of CENTRIC

EXECUTIVE SUMMARY

One of the challenges for internal security practitioners, and particularly law enforcement agencies (LEA), is to determine how to capitalise on Artificial Intelligence (AI) capabilities in response to changing safety and security challenges while, at the same time, ensuring true accountability of its use. Accountability is not only of societal interest; in the internal security domain it is often also a legal requirement. Yet, accountability is easily claimed but hard to evidence.

The AP4AI Project addresses this challenge by creating an AI Accountability Framework which will allow practitioners to capitalise on AI capabilities, whilst demonstrating meaningful accountability towards society and oversight bodies. The AP4AI Framework will be a practice-oriented mechanism designed specifically for internal security practitioners to proactively assess, as well as reactively demonstrate AI Accountability.

The AP4AI Project started in 2021 and is jointly coordinated by CENTRIC and Europol and supported by Eurojust, EUAA and CEPOL, with advice and contributions by FRA in the framework of the EU Innovation Hub for EU Internal Security.

This current report summarises the overall project approach, as well as an overview of the outcomes of the first project phase which stem from consultations with subject matter experts from law enforcement agencies and border police, justice and judiciary, human rights, law, ethics, civil society, NGOs, industry and academia in 28 countries. These consultations delivered 12 core Accountability Principles for AI for the internal security domain which are described in this document. These 12 initial principles lay the foundation for the further development of the AP4AI Framework in the coming months, including concrete implementation guidelines and AP4AI assessment toolkits.

By integrating international perspectives by security, legal, ethical and technical communities as well as citizens on use of AI in the internal security domain, the AP4AI Framework will lead to a step-change in the application of AI by the internal security community.

CONTENTS

INTRODUCTION

Artificial Intelligence (AI)[4] is a critical asset for the effectiveness and efficiency of the internal security community, including law enforcement and the justice sector.[5] Security is an information-based activity, for which AI applications can provide crucial support in all steps from acquisition to analysis, decision support and the collection of evidence. AI can thus create important resource efficiencies and performance gains; for example, by optimizing the evidence gathering and analysis process in serious and organised crime cases or by aiding the discovery of new adversarial trends and malicious patterns. Simultaneously, citizens as well as security practitioners themselves raise legitimate concerns, chief amongst them that AI use can reinforce social inequalities, lead to faulty decisions with dramatic real-life consequences and create inflexible, insensitive procedures that fail to take into account individuals' unique circumstances yet cannot be challenged because the underlying rules are too complex or opaque.[6]

Given the assumption that law enforcement and criminal justice institutions are able to operate effectively only to the extent that they are entrusted by society with the restriction of personal freedoms for the pursuit of enforcing the law and the provision of safety and security, the use of algorithms and AI-based systems and platforms must be not only carefully understood but also be accountably scrutinised by and responsive to the relevant public and oversight authorities. For all organisations which have security and justice as a core mandate, accountability is thus essential in ensuring a successful relationship with citizens. In fact, in many instances establishing the necessary arrangements for democratic accountability is in fact a legal requirement.[7]

The challenge for internal security practitioners including law enforcement and the justice sector is to determine how to capitalise on the opportunities offered by AI to improve the way investigators, prosecutors, judges or border guards carry out their mission of rendering justice and keeping citizens safe, while at the same time safeguarding and demonstrating true accountability of AI use towards society.

The AP4AI (Accountability Principles for Artificial Intelligence) Project addresses this challenge by creating a comprehensive and validated **Framework for AI Accountability for Policing, Security and Justice**. With this, AP4AI offers a step-change in the application of AI by the internal security community by defining a robust and application-focused Framework that integrates security, legal, ethical as well as citizens' perspectives on use of AI in the internal security domain, including law enforcement agencies (LEAs) and the justice sector.

AP4AI: ACCOUNTABILITY PRINCIPLES FOR ARTIFICIAL INTELLIGENCE

The AP4AI Project develops solutions to help research, design, assess, review and revise AI-led applications in a way that is both internally consistent and externally compatible with the respective jurisdictions of widely differing organisations, while safeguarding accountability in AI usage by practitioners in line with EU values and fundamental rights. To this end, AP4AI will create a Framework for security and justice practitioners including law enforcement agencies (LEAs) which integrates central indefeasible tenets which, if adopted, will provide practitioners, legal and ethical experts as well as citizens with a degree of reassurance and redress. The AP4AI Framework will allow practitioners to capitalise on available AI capabilities, whilst demonstrating meaningful accountability towards society and oversight bodies. The AP4AI Framework will provide a mechanism to proactively assess, as well as reactively demonstrate AI Accountability.

The project's ambition is driven by the recognition that enabling technologies in sensitive and complex areas such as offender profiling, identification of illegal internet content (e.g. child sexual exploitation materials), live facial recognition (LFR), inferential algorithms, 'new biometrics' such as gait analysis, robotics, cyber security, ambient intelligence (AmI), Internet of Things (IoT) and nanotechnology raise a new generation of considerations such as multiple identities, jurisdiction, intellectual property in genealogical biometric data and proof of a 'controlling mind' which will need to be calibrated and balanced against and adhere to existing individual rights and legitimate expectations.

AP4AI focuses on accountability as guiding standard, under the premise that in the field of security and justice, AI Accountability is as important as the technology itself.

AP4AI's accountability perspective is based on the understanding that the extent to which security practitioners are *accountable* to their communities is a proxy measure for the extent of their *legitimacy* within those communities. Rather than proposing a further fixed set of rules as an addendum to the formal legal and regulatory frameworks that are already applicable within their jurisdictions, the AP4AI Project offers a fundamental set of inter-connected and citizen-validated principles for:

(a) internal community practitioners and their partners to demonstrate their accountability when designing, (de)commissioning, procuring and utilising AI and
(b) oversight bodies and the public to measure security practitioners' use of AI against.[8]

In this way, AP4AI seeks not only to guard against *misuse* of AI, but also to *ensure accountability in a broader sense* across all phases and aspects of AI use and application by law enforcement agencies, justice agencies and their partners whichever domestic jurisdiction they operate within.

The AP4AI Accountability Principles, on which the AP4AI Framework are based, offer an applied mechanism to assess and enforce legitimate and acceptable usage of AI by the internal security community and are intended to guide the research, design and application of current and future AI capabilities within the security and justice domain. The AP4AI Principles are thus intended for internal security practitioners to demonstrate accountability in a way that can be tested by presenting available evidence against a carefully researched and accessible standard.

Moreover, the AP4AI Principles can guide and inform legislation bodies to create future-proofing legislation and enforcement directives agnostic of particular technological changes. This is a step-change in the application of AI in law enforcement, as it offers a unified, consistent and practical mechanism for security practitioners, oversight bodies and the public to ensure enduring AI accountability.

This report describes the initial set of 12 accountability principles which result from AP4AI's first project phase, namely the intense engagement process with AI experts in 28 countries covering 22 EU Member States, Australia, Canada, Norway, Ukraine, UK and USA. It further summarises the expert opinions and recommendations with respect to AI accountability from the diverse set of experts consulted in this first phase and outlines the subsequent steps in AP4AI towards its objectives.[9]

The AP4AI project identifies Fundamental Rights as a key element and a mandatory requirement of the legal framework governing the use of AI for the internal security community. In the next iterations of this report, AP4AI will address Fundamental Rights in relation to use of AI by internal security practitioners. The latter will cover applicable standards and research on AI and fundamental rights, including that conducted by the EU Agency for Fundamental Rights (FRA)[10] and others. The project will also review existing work on impact assessments for the use of AI[11] to inform the AP4AI implementation guidelines and toolkit.

ACCOUNTABILITY AS GUIDELINE FOR AI USE BY LEAS AND THE INTERNAL SECURITY ECOSYSTEM

The AP4AI approach[12] uses accountability as the core guiding value for AI deployments in the internal security domain. Accountability is intended for "preventing and redressing abuses of power".[13] Following this concept, AP4AI understands accountability as the responsibility to fulfil obligations towards one or multiple stakeholders, in the understanding that not meeting these obligations will lead to consequences. *AI Accountability* translates this concept to the AI domain encompassing AI users (e.g., police organisations) and deployments (e.g., systems, software platforms, usage situations).

Accountability comprises in itself the three aspects of monitoring, justification and enforcement,[14] and in a legal perspective is defined as the "acknowledgement and assumption of responsibility for actions, decisions, and their consequences."[15] It thus has at its very core the notion of negotiation across disparate legitimate interests, the observation of action and consequences and the possibility for redress.

Accountability is the acknowledgement of an organisation's responsibility to act in accordance with the legitimate expectations of stakeholders and the acceptance of the consequences – legal or otherwise – if they fail to do so. In this context liability or rather 'answerability'[16] is the basis for meaningful accountability as it creates a foundation for the creators and users of AI to ensure that their products are not only legally fit for the legitimate purpose(s) in the pursuit of which they are used (attracting the appropriate claims for negligence or other breach of duty as fixed in law), but also invite scrutiny and challenge and accept the consequences of using AI in ways that their communities find morally or ethically unacceptable. There is further responsibility to ensure the avoidance of misuse and malicious activity in whatever form by both the relevant security practitioners and their contractors, partners and agents. AP4AI, by focusing on AI Accountability, is a framework designed to underscore the importance of legal, ethical and societal duties of responsible organisations using AI in a security context, which explicitly includes consequence for misuse and breaches in conduct.

Included in the above is the crucial aspect of public scrutiny, which sustains the maintenance of public confidence in the authorities' adherence to the rule of law and the prevention of any appearance of collusion in, or tolerance of unlawful acts. It further encompasses organisational accountability as precondition for AI deployments that are responsive and responsible within the organisations themselves as well as towards outside oversight structures.

We argue for the primacy of accountability as guiding framework for AI use in the internal security domain as it is the only concept that binds organisations to citizen-enforceable obligations and thus provides a foundation that has actionable procedures at its core. The AP4AI Framework will create a practical mechanism (including an assessment toolkit) to ensure that legitimate interests (as well as concerns, fears and expectations) of all stakeholders are engaged with and factored in throughout the full decision-making process about security-related AI capabilities and use.

The notion of accountability offers vital benefits compared to other instruments and frameworks. Many existing instruments and frameworks offer descriptive labels or desired end-states such as "Responsible AI" or "Ethical AI" which, while useful to inform, aspire and educate stakeholders within the security, policing and justice sector, tend to be limited in their scope and practical implementation partly because they are designed by and for 'experts' and do not empower the citizen with the knowledge to hold their institutions accountable for their use of AI. Most importantly, they lack the element of recognising the need for answerability in this sector, which apprehends enforcement in fulfilling all relevant *legal* obligations as well as the acceptance of material consequences should they fail to meet the legitimate expectations of their governance bodies. Without either element an instrument or framework is one-sided; without both they cannot be considered as offering a holistic governing framework for AI usage within the internal security domain.[17]

Yet, while 'Accountable AI' is an established concern and aspiration for AI system designers,[18] there is a profound AI accountability gap with respect to societal, organisational, legal and ethical answerability by which to understand and sustainably manage the complexities of AI in the security domain in a way that affords adequate monitoring *and* enforcement. Also, disturbingly, there is little clarity on what the generic term accountability means in the complex societal, organisational, legal and ethical sense in the context of security-related AI applications. For instance, while legal and procedural accountability in general is a well-established concept for LEAs, at present there is no reliable definition of accountability in the context of AI applications for the internal security domain.

Even from a legal perspective very few cases refer to the use of regulatory powers from data protection authorities to clarify the boundaries of security practitioners' accountability in the use of AI,[19] certainly in pan-European legal interpretations. While there is a body of evidence corroborating the concerns about the potential impact of AI in this sector – particularly in the Visual Surveillance of Populations[20] – there is currently also no clear legal definition of 'accountability' in the EU jurisprudence. It also remains to be defined how accountability interrelate throughout the process of an AI system's lifecycle including the development of disparate AI tools, applications and platforms for security practitioners, their usage and decommissioning/replacement.

In AP4AI, accountability is approached as a relational concept in that obligations are directed towards and between particular stakeholders or groups. In a law enforcement or security context, discussions of accountability tend to be focused on police accountability towards citizens. Given the complexity and the scale of effects security applications of AI have on individuals, communities, societies and organisations (LEAs and others) not only at local, national, and European levels but increasingly at a global level[21] this is insufficient.

Instead AP4AI work is informed by the conviction that all AI stakeholders (citizens, security practitioners, judiciary, policy makers, industry, academia, etc.) have to be active constituents in the accountability process, and that this process needs to be grounded in a broad and sustained engagement.[22]

The innovative potential of AP4AI is in establishing the extent, form and nature of accountability in relation to *society* (including needs and legitimate expectations of individuals and specific groups), *LEA and internal security organisations, law and ethics*, and their translation into (a) overarching, universal principles to guide current and future AI-capabilities for the internal security community guided by EU values and fundamental human rights and (b) the conception of methods and instruments for their context-sensitive and adaptive implementation.

AP4AI OBJECTIVES

The AP4AI Project aims to represent the very specific accountability requirements use in the internal security domain suitably adapted to meet today's realities of AI deployment in the law enforcement and internal security ecosystem including the EU internal security community and justice sector.

In a first step AP4AI has defined a novel set of inter-connected, operationally focused, implementable guiding principles for the internal security domain that cover the full AI lifecycle from ideation to application and potential adaptation based on human, societal, organisational, legal and ethical assessments founded on the notion of police accountability.[23] The AP4AI Principles, and the AP4AI approach in general, can be considered a further application of the Peelian Principles[24] and are built on the Principles of Accountable Policing proposed by the Scottish Universities Insight Institute in 2016.[25]

AP4AI incorporates a formal public expression of commitment by relevant office holders against which they may be held directly and publicly to account. AP4AI thus emphasises sustained and broad engagement around AI needs, mechanisms and procedures within the internal security community that can be used efficiently, are transparent and fair, adaptive across contexts and thus able to become an established part of any AI research, development and deployment efforts of internal security practitioners on local, national and European levels.

To be productive, the AP4AI accountability principles need translation into actionable steps. The project will thus formalise an **Accountability Principles for AI (AP4AI) Framework** in support of Freedom, Security and Justice (to be published later this year).

In the longer-term it is AP4AI's ambition to provide an evidenced-based, comprehensive methodology for the implementation and adaptation of the AP4AI Framework for the disparate stakeholder groups relevant to the AI Ecosystem (i.e., LEA organisations, citizens, data protection officers, local or national policy makers, technology providers, researchers, etc.).

PROJECT METHODOLOGY

This section summarises the general approach within the AP4AI Project, as well as the empirical work leading up to the current report.

OVERALL APPROACH

To ensure the robust development and validation of Accountability Principles for AI, the project is conducted in three cycles that employ a sequential mixed method approach with consecutive steps of exploration, integration and validation:

- **Cycle 1 – Development of the AP4AI Principles (ongoing)**: The first cycle consisted of two activities: (a) a comprehensive review of existing frameworks aiming to guide or regulate AI and (b) expert consultations with subject-matter experts from law enforcement, justice, legal, ethical and technical fields identified by the AP4AI consortium partners
- **Cycle 2 – Citizen consultation for validation and refinement of the Principles (ongoing)**: An online survey with approximately 6,000 citizens in 30 countries including all EU members states, UK, USA and Australia will collect citizen input on the AI Accountability Principles developed in Cycle 1.[26] The citizen perspectives will be integrated with Cycle 1 results leading to the preliminary AP4AI Framework.
- **Cycle 3 – Expert consultation for validation and completion of the AP4AI Framework (upcoming**): The preliminary Framework will be sent to subject matter experts from Cycle 1 and new experts invited for review and validation. The mixture of existing and new subject matter experts will ensure that (a) experts familiar with the past work can comment on the treatment and coverage of past inputs and (b) new experts unfamiliar with past work can independently verify outcomes and potentially supplement additional aspects. The results of this consultation will be consolidated into the final AP4AI Framework.

Engagement with a broad range of stakeholder groups is key to AP4AI to ensure that the AP4AI Framework is grounded in a comprehensive understanding of Accountability and Artificial Intelligence in the internal security domain. AP4AI consults and engages with subject matter experts from the following stakeholder groups:

1. AI experts from law enforcement agencies and border police
2. Justice and Judiciary
3. Human rights experts
4. Legal AI experts
5. Ethical AI experts
6. Civil Society and NGOs
7. Technical AI experts

Most importantly, the project consults and engages in Cycle 2 with the principal group in any democratic policing and justice model: **the citizen**. If the citizen in whose name these functions purport to be done – and at whose expense – is not involved centrally and meaningfully, any framework claiming to enhance democratic accountability lacks structural credibility.

The AP4AI consortium further recognises that AI use in the internal security domain – whether at practitioner or citizen level – is strongly affected by the national contexts in which AI capabilities are deployed. The consortium therefore conducts its consultations across 30 countries: all 27 EU countries, UK, USA and Australia.

DEVELOPMENT OF THE AP4AI PRINCIPLES (CYCLE 1)

The objective of Cycle 1 is to develop a validated set of universal AI Accountability Principles for the internal security domain, while also investigating potential differences amongst stakeholder groups in their perspectives on AI Accountability. Cycle 1 comprised of two activities:

1. A comprehensive review of existing AI frameworks, guides and policy statements published by national and international organisations from 2017
2. Subject matter expert consultations with AI experts from all seven stakeholder groups listed above

Review of existing AI frameworks, guides and policy statements

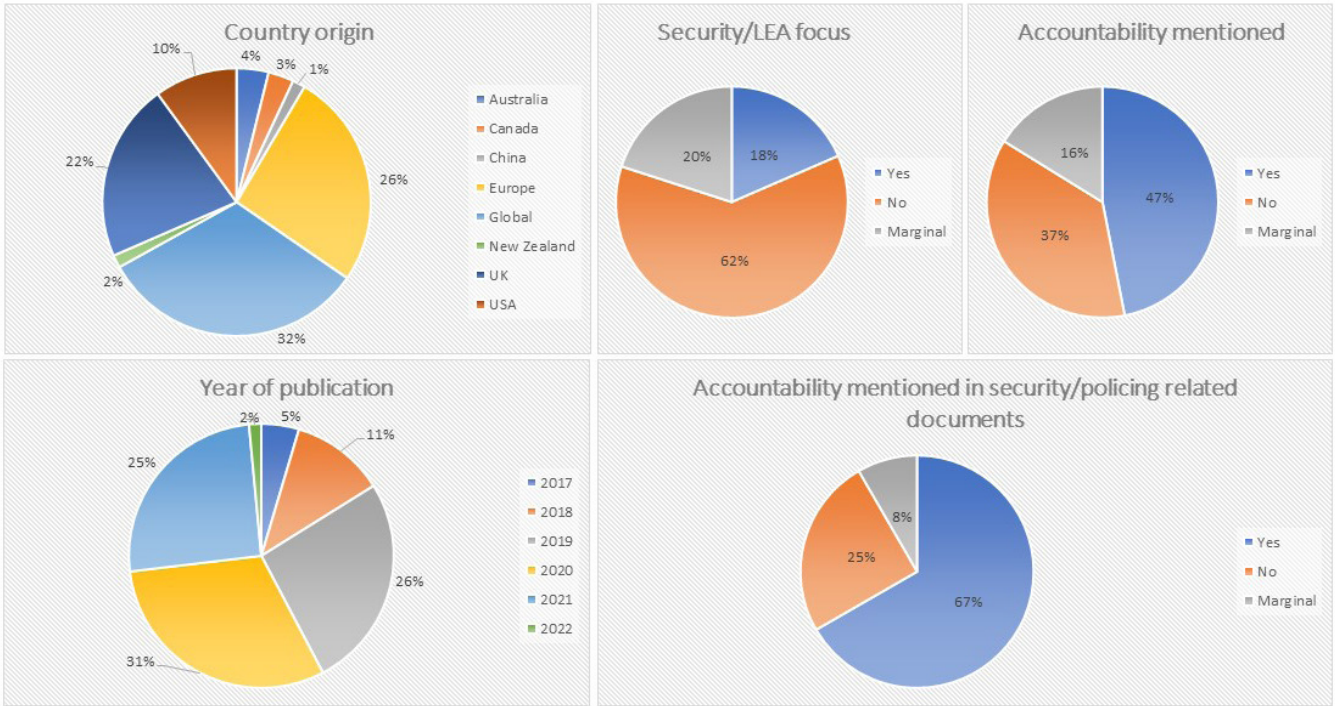
To ensure that AP4AI work and results are cognisant of as well as able to relate to and reflect latest developments, a comprehensive review of existing documents and reports was conducted. The selection of documents was purposefully broad to guarantee an expansive search. The following criteria were applied:

- *Inclusion criteria*: document has AI as core topic, document is publicly available, publication date is 2017 or later, any type of publication (reports, articles, white papers, chapters, etc.), any type of publishing organisation (national body, international body, public organisation, private company, academia, NGO, etc.)
- *Exclusion criteria*: published before 2017, AI is only addressed in passing (e.g., as example), document not in English

Overall, 130 relevant documents were identified until November 2021. Documents were analysed using a standardised coding scheme with the following categories for (a) *meta-information*: documents addresses accountability (yes, no, marginally), is focused on security/law enforcement domain (yes, no, marginally), mentions specific principles related to the use of AI (yes, no), discusses citizen perspectives (yes, no, marginally); (b) *content*: accountability definitions, type of principle(s) addressed, sections that addressed any of the 14 policing principles used as starting point for the investigation (see Table 1 in section *Collection of pre-consultation input* for an overview of the principles).

Figure 1 provides a summary of the most relevant meta-information. As the summary illustrates, the majority of the relevant documents were published in 2020 and 2022 (57.7%), while the focus was primarily on the European context (26%; e.g., publications by European Commission), global/international considerations (32%; e.g., OECD), UK (22%) or USA (10%). Only a small percentage had a clear security/law enforcement focus (18%), compared to 62% without any mention of security or policing. Accountability was mentioned as a consideration for AI in 47% of reviewed documents.[27] This number increased to 67% for security-related documents demonstrating the relevance of accountability for this area. However, none of the reviewed security/law enforcement related documents focused exclusively on accountability or aimed to define accountability and its component mechanisms for AI usage by LEAs.[28]

Figure 1: Summary of relevant meta-information of the reviewed documents

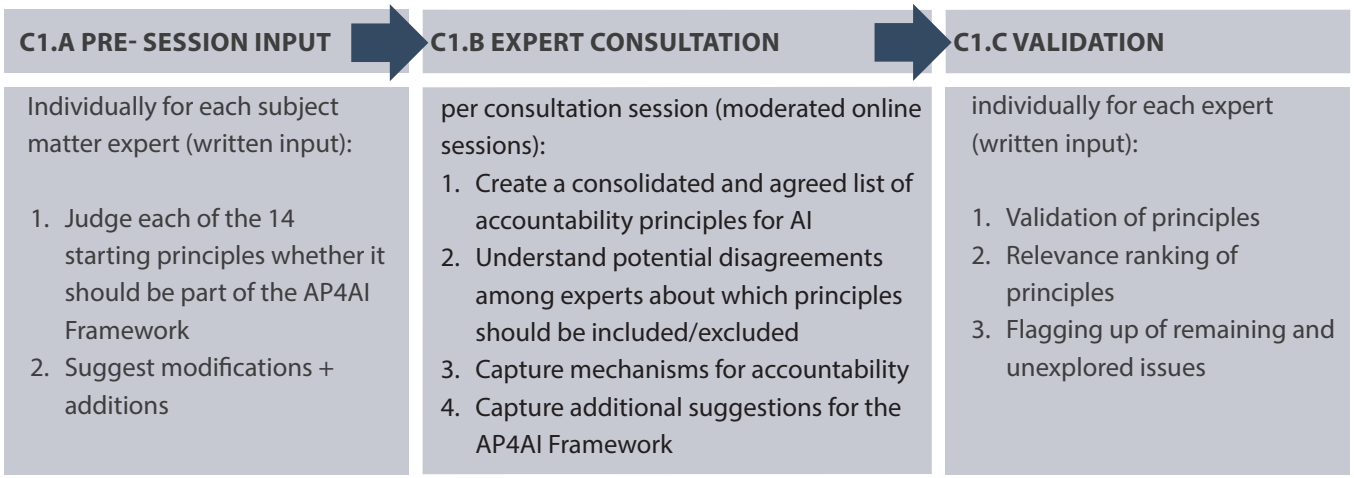


Subject matter expert consultations

The subject matter expert consultations comprised of three steps:

a. *Collection of written pre-consultation input (completed)*: Experts were asked to provide their assessment of 14 general principles in written form as well as list additional principles deemed missing in a structured template
b. *Expert consultation session (completed)*: Consultation sessions were moderated focus group discussions to reflect on inputs in a group of experts with the same disciplinary background (i.e., law enforcement, legal/ethical expertise, technical expertise). The objective was to obtain an agreed list of accountability principles for AI, understand potential disagreements among experts about which principles should be included/excluded, as well as reflections on the AP4AI approach generally. The consultation sessions were recorded and transcribed verbatim. For experts unable to attend a consultation session, only the written input was collected using the same template as for the pre-consultation input.
c. *Validation of core principles (ongoing)*: Experts who participated in the consultation sessions will receive a summary of the consolidated expert inputs for comment and validation using structured validations forms.

Figure 2: Steps conducted in Cycle 1



Collection of written pre-consultation input

The written pre-consultation input was collected using a structured template. The structured format guaranteed that inputs were focused, easy to compare and easy to integrate across participants and reduce the time commitment on participating AI experts. The starting point for the consultation were the 13 law enforcement agency principles of good practice proposed by Fyfe et al. (2020)[29] plus the principle of Trustworthy AI put forward by the European Commission's High-Level Expert Group on AI[30]. Apart from Trustworthy AI, these principles are not AI specific. However, they represent a rare set of established accountability norms for the law enforcement domain and thus constituted a legitimate starting point for discussions about accountability in the much more targeted and practical area of AI deployment in the internal security domain. Table 1 provides an overview of the 14 starting principles as well as simplified definitions.

Table 1: Overview of the 14 principles as starting point for the expert consultations

1. **Universality**: requires that all relevant manifestations of AI in policing are in scope, including contractors and technology providers carrying out functions on behalf of LEAs.
2. **Independence**: requires bodies responsible for holding the police to account for the development and deployment of AI to demonstrate how they are sufficiently distinct from policing in order to enhance public trust and confidence.
3. **Compellability**: an effective accountability AI regime must afford an independent accountability body the capacity, capability, authority and opportunity to interrupt, interrogate and, if necessary, compel.
4. **Enforceability and redress**: requires that citizens who believe they have been wronged by the LEA's use of AI have an accessible and meaningful avenue of redress.
5. **Legality**: ensures that LEAs' use of AI is subject to the same strictures and consequences of misconduct as would apply to any other person.
6. **Conduct**: follows the international legal framework and incorporates elements of effective investigation of police complaints[31] and promotes the relevant standards and behaviours and facilitate complaints and compliments.
7. **Constructiveness**: requires LEAs to make clear how and why to complain and to assign sufficient resources to complaints, assuring that someone will listen, that something will be done and that something will change.
8. **Clarity**: aims to establish a shared understanding amongst all stakeholders in the AI project's lifecycle.
9. **Transparency**: includes the availability and ready accessibility of relevant information and datasets (so far as is appropriate and by consideration of legitimate security and operational needs of LEAs).
10. **Pluralism and Multi-level Participation**: posits that, if claims to the 'public good' are to be made for AI, then the public has to be engaged throughout the accountability processes, taking also careful account of the historical challenges in involving marginalised groups.[32]
11. **Recognition and Reason**: aims to facilitate 'participatory space' and encourage authentic public scrutiny.[33]
12. **Commitment to Robust Evidence and Independent Evaluation**: recognises that deliberations need to be informed by robust evidence and rigorous, independent evaluation of outcomes.
13. **Be a Learning Organisation**: requires embedded formal systems to ensure that lessons are learnt from incidents and errors openly and systematically.
14. **Trustworthy AI**: AI systems need to be based on the principle of trustworthiness, i.e., be lawful, ethical and robust.

Experts were asked to provide their assessment for each of the 14 principles on whether to: (a) include the principle as is, (b) include the principle with adaptations or (c) not include the principle. If experts chose options B or C, they were asked to provide a description of the change or justification for the deletion (see Figure 3 for an illustration). They were further asked to add AI accountability principles they thought were missing.

Figure 3: Excerpt of the pre-consultation template

Principle *(listed in random order)*	**A:** **Include as is**	**B:** **Include, but needs adapta-tion**	**C:** **Do not include**	**Explanations** **If B: explain adapta-tion** **If C: explain why**
Universality all relevant manifestations of policing should be in scope including external contractors and tech partners processing data or carrying out functions on behalf of LEAs	☐	☐	☐	

Expert consultation sessions

The consultation sessions were organised as discipline-specific discussions (i.e., with participants in one session stemming from the same stakeholder group, although representing different countries). The choice for using homogeneous – in preference to mixed – stakeholder groups was made to facilitate in-depth and detailed discussions on specific, often discipline-specific issues (e.g., laws or operational police challenges) which experts may not be willing or able to share with people outside of their profession.

The discussions were guided by the results of the pre-consultation inputs in that written input by participants in the same session was summarised to showcase agreements/disagreements in opinions for each of the 14 starting principles. Summarising the inputs led to three groups: (a) principles all experts in the session agreed should be kept as is, (b) principles the majority of experts in the group suggested should be kept but adapted, (c) principles with strong disagreements in the group, i.e., with experts' opinions ranging from 'keep-as-is' to 'delete' for the same principle. The moderated discussions investigated the reasons for deletion and adaptation decisions as well as reasons for differences in judgements. Lastly, additional principles proposed by individual experts were reviewed within the group to obtain a broader opinion on the AI4AI Framework.

All sessions took place online to facilitate participation of subject matter experts from a large range of countries and to eliminate burdens on experts' time.[34] Session length was capped at 2 hours. All expert consultation sessions were recorded and transcribed verbatim to allow detailed content analysis.

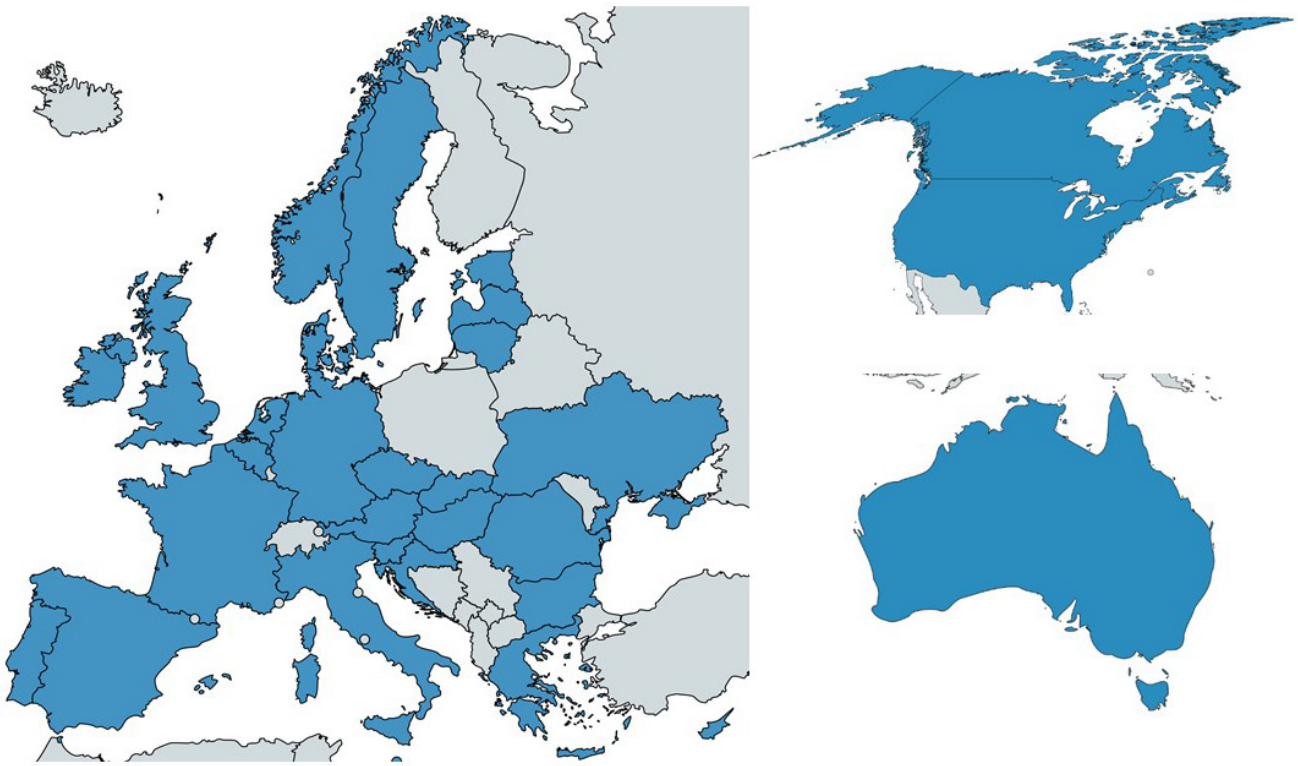
Exert inputs collected

Overall, inputs from 69 subject matter experts were collected in Cycle 1. Of these, 49 were from law enforcement agencies, eight from technical experts, seven from legal experts and five from ethical and civil society experts. As part of these engagements, six expert consultation sessions took place – three with experts from law enforcement agencies, one with technical and legal experts and one with ethics and civil society experts:

- **08/04/2021: Expert domain**: Legal; **Participants**: Public prosecutor, Prosecutor, Judges, liaison prosecutor, Justice sector experts
- **04/05/2021: Expert domain**: Law enforcement; **Participants**: Interior ministries, counter-terrorism experts, national police forces
- **05/05/2021: Expert domain**: Technical; **Participants**: Private sector AI providers, Software developers, Academia (Technical)
- **02/06/2021: Expert domain**: Human rights; **Participants**: Fundamental Rights, NGOs, Academia
- **17/06/2021: Expert domain**: Legal; **Participants**: Academia (Law)
- 1**4/07/2021: Expert domain**: Law enforcement; **Participants**: Law enforcement agencies

In accordance with the ambition for a broad, international consultation, the inputs cover 28 countries (22 EU Member States, Australia, Canada, Norway, Ukraine, UK and USA), as well as input from experts in multinational organisations (e.g., Europol, FRA, Eurojust, EUAA, societal organisations with European or global reach). Figure 4 indicates the countries with participation in Cycle 1.

Figure 4: Countries in which experts were located



Analysis of inputs

The development of the AP4AI Principles followed a 3-step process: (a) coding of inputs, (b) consolidation of information from multiple coders, (c) selection into the final set.

Coding of inputs: The session transcripts and written pre-consultation inputs were analysed by a team of four researchers. Using thematic coding, the content was coded along seven core themes: (a) type of changes requested per principle; (b) reasons for deletion of a principle or alternatively (c) whether a principle was marked as 'keep-as-is'; (d) comments on the AP4AI Framework overall; (e) comments on the concept of accountability; (f) organisations or actors that should be involved in or responsible for the accountability process and (g) principles suggested by experts in addition to the 14 principles originally proposed.

Consolidation of data by multiple coders: Coded information for each of the 14 principles was analysed independently by two of the four coders and counterchecked against information in existing AI frameworks. Integration sessions between the two researchers provided a consolidation per principle as well as a view on potential overlaps between principles.

Selection into the final set was achieved in a common review of all evidence by the four coders, accompanied by the moderator of the expert sessions. Selection of the principles was guided by two considerations: (a) retaining as broad a perspective on AI Accountability as possible accommodating the diverse professional perspectives across stakeholder groups and (b) reducing overlaps amongst principles to ensure each principle addresses a unique aspect of AI Accountability.

Experts collectively made 34 suggestions for additional principles or for the rephrasing of the initial 14 principles. The list of suggestions can be found in Table 2. The suggestions were carefully reviewed and compared to the initial 14 principles. A number of suggestions provided important additions and elucidations for existing principles. Such suggestions were included in the content of the respective principle (e.g., 'learning from accountability process itself' which is a crucial element for the principle of Learning Organisation). Where this was the case, Table 2 marks them as 'addressed in', indicating that this aspect was added to the respective principle. Other suggestions addressed important mechanisms to ensure accountability (marked as 'mechanism' in Table 2). These suggestions will form a vital part in the further development of the AP4AI Framework, which will also consider possible mechanisms for the practical implementation of the AP4AI Principles.

Table 2: List of additional principles and aspects suggested by experts	
Impartiality to avoid conflicts of interest	Addressed in: Constructiveness
Welcoming oversight	Addressed in: Constructiveness
AI requires transparent + understandable outputs	Addressed in: Transparency
Open data	Addressed in: Transparency
Non-recursive transfer operational data	Addressed in: Transparency
Human right impact assessment before purchase, deployment	Addressed in: Legality (as mechanism)
Human rights	Addressed in: Legality
Privacy + data governance	Addressed in: Legality
Procedural rights	Addressed in: Legality
Confidentiality, data protection	Addressed in: Legality
Demonstrable data protection	Addressed in: Legality
Need to use advanced technologies to protect human rights	Addressed in: Legality
Proportionality with respect to AI system criticality	Addressed in: Legality
Data governance	Addressed in: Legality
Worker autonomy + responsibility	Addressed in: Learning Organisation
Learning from accountability process itself	Addressed in: Learning organisation
Auditability	Addressed in: Commitment to robust evidence
Scientific robustness	Addressed in: Commitment to robust evidence
Technical robustness + safety	Addressed in: Commitment to robust evidence
Awareness of abuse	Addressed in: Enforcement and Redress
Good administration of AI	Mechanism
Certification	Mechanism
Certification of oversight bodies	Mechanism
Declaration regime (audits, etc)	Mechanism
Evaluation of tools before, after use	Mechanism
Regime of sanctions	Mechanism
Regular evaluation	Mechanism
Human oversight	Mechanism
Trustworthy LEA	Overall ambition rather than a principle
AI that is specific for systems trained and used in LE context	Overall ambition rather than a principle
Addressing the pacing problem, fast development of AI	Overall challenge rather than a principle
Non-use of AI must be a viable outcome	Overall challenge rather than a principle
Explainability	Added as separate principle

Final set: Of the 14 initial principles 11 principles were retained (see section *Description of AP4AI Principles*). From the additional principles suggested by experts we included Explainability as a twelfth principle, as it was named consistently as a crucial standard for accountability.

Additional insights: Next to informing the initial set of Accountability Principles, the expert consultations also highlighted important considerations for the further development of the AP4AI Framework. These considerations address the presentation of the Framework, the role of fundamental rights and national laws, mechanisms to assure accountability, clarification of possible exceptions and groups relevant for AI accountability in the internal security domain. A summary of core insights can be found in Appendix A. These considerations have been reviewed in detail and inform upcoming activities in the AP4AI Project (see Appendix A and section on *Upcoming activities*).

DESCRIPTION OF AP4AI PRINCIPLES

This section provides the high-level outline of the AI Accountability Principles using a uniform structure and where necessary additional elaboration and examples. Each principle is explained individually describing its meaning, practical considerations for its implementation, aspects of note and examples of applicable laws.

The aim of this section is to present each principle in a concise form to convey the 'essence' of the AP4AI approach. This is intended to make the Principles more accessible at this early stage, allowing different stakeholders to consider their application in a specific context and determine their relevance for their specific requirements. This report thus provides an overview rather than a detailed discussion and implementation guideline. The latter will be provided in subsequent reports of the AP4AI Project.[35]

Table 3: Initial set of AP4AI Principles

1. Legality	7. Enforceability
2. Universality	8. Compellability
3. Pluralism	9. Explainability
4. Transparency	10. Constructiveness
5. Independence	11. Conduct
6. Commitment to Robust Evidence	12. Learning Organisation

1. LEGALITY

Meaning

All aspects of the use of AI should be lawful and governed by formal, promulgated rules. This may seem axiomatic but the starting point for Accountability requires that compliance with applicable international, national and sector-specific laws, rules, norms and agreements should be clearly identified and demonstrated. In addition to the core aim of mitigating risks to fundamental rights and freedoms, the principle of Legality extends to all those involved in building, developing and operating AI systems for use in a criminal justice context. Where any gaps in the law exist, the protection and promotion of fundamental rights and freedoms should prevail.

Practical Considerations

- How do the applicable laws apply in this context?
- How can compliance be demonstrated?
- Are the overriding principles of necessity and proportionality complied with?
- Do any legal exemptions apply? If so, are appropriate safeguards in place?
- Is the appropriate oversight body engaged, in respect of the activity?
- Despite legal compliance, any residual risks particular to AI should be addressed.
- Legal compliance alone may not address wider public concerns.
- Some aspects of AI usage, including new developments and capabilities, may not be regulated in existing laws and standards.

Of Note

It is envisaged that Algorithmic Impact Assessments (AIAs) will play an important part in the implementation of this principle and are aligned with the approach set out in the EU's proposed Artificial Intelligence Act.[36] The use of AI regulatory sandboxes is also promoted in the proposed Act, which will play an important part in identifying risks and potential consequences, as well as measures needed to achieve legal compliance, in a safe environment.

Elaboration - Examples for supporting mechanisms:

"Member States will be encouraged to launch AI regulatory sandboxes to promote the safe testing and adoption of AI systems under the direct guidance and supervision of national competent authorities. [….] A new European AI Board will be established to facilitate the consistent implementation of the regulation, comprising representatives from Member States' national competent authorities, the European Data Protection Supervisor, and the Commission" (EU Artificial Intelligence Act)

AI Impact Assessments (AIAs) could help determine the impact of an AI deployment and establish whether an AI system will break EU laws including human rights and GDPR principles. Excerpt from expert input: An AIA "gives both the agency and the public the opportunity to evaluate the adoption of an automated decision

system before the agency has committed to its use. This allows the agency and the public to identify concerns that may need to be negotiated or otherwise addressed before a contract is signed. This is also when the public and elected officials can push back against deployment before potential harms occur." To note, AIAs follow a similar format to measures suggested to ensure the ethical implementation of 'high-risk AI system' as specified in the European Commission's Artificial Intelligence Act.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect[37].
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- National industry or sector-specific legal standards relating to public safety.
- National and sector-specific tribunals and formal procedures providing means of effective redress in applicable contexts.

2. UNIVERSALITY

Meaning

Universality provides that all relevant aspects of AI deployments within the internal security community are covered through the accountability process. Effectively extending the 'jurisdiction' of the principles to all who are subject to the legality principle (above), this principle recognises the reality that AI applications are necessarily multi-partner input programmes in a frequently complex process and the need for public trust and confidence must extend to the whole ecosystem. This is not only in respect of the deployment of AI in a criminal justice context, but in all the related processes, including design, development and supply, to which accountability applies equally (including all domains, aspects of police mission, AI systems, stages in the AI lifecycle or usage purposes), and prevents contracting out or off-shoring by the relevant accountable organisation.

Practical Considerations

- This principle applies to all components and the complete life-cycle of an AI system, from design to decommissioning/replacement. How has this been mapped out?
- Do all those involved understand their responsibilities in respect of compliance with accountability and therefore this principle? How is this ensured?
- How is compliance with this principle measured? Who is responsible for this?
- Have all processes affected by AI been accounted for?
- Have all relevant stakeholders been considered?
- Have efforts been made to understand concerns relating to specific sections of society, as well as the wider public? How will these be addressed through compliance with this principle?
- Have all outcomes and possible impacts of AI deployment been considered?
- Have all aspects of oversight bodies and mechanisms been considered?

Of Note

There may be restrictions in achieving Universality, for example, due to legal or sector-specific constraints in respect of types of information. In the name of accountability, any restrictions should be recorded in a specific and clear way, including justifications and mitigating measures adopted in respect of achieving accountability.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.

- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- National industry or sector-specific legal standards relating to public safety.
- National and sector-specific tribunals and formal procedures providing means of effective redress in applicable contexts.

3. PLURALISM

Meaning

Pluralism ensures that oversight involves all relevant stakeholders engaged in and affected by a specific AI deployment. Pluralism avoids homogeneity, where all those regulating seem to come from the same background as those who are being regulated and thus a tendency or perception for the regulators to take a one-sided approach. Participation should be achieved through a combination of democratic processes and consultative forums at national and local levels.[38]

Practical Considerations

- Is the selection of stakeholders sufficiently comprehensive?
- Has the adequate (local, national, cross-national) level of participants been achieved?
- How have stakeholders or interested parties been identified or defined?
- Have a variety of methods of engagement been implemented, in the true spirit of inclusiveness?
- Has full information about procedures been provided in a clear and meaningful way, which also achieves the management of expectations?
- Do participants understand their role within the process and the purpose of it?
- In which form should law enforcement agencies be included?
- What form should citizen engagement take?
- Should stakeholders remain the same at all stages of accountability procedures and engagements?

Of Note

Awareness must be maintained of considerable challenges to be overcome or accounted for in respect of this principle. In particular, reluctance to engage or perceptions of misalignments between rhetoric and reality. In relation to methods of engagement, it is anticipated that remote contribution capabilities have significantly improved as a consequence of the COVID-19 pandemic.

Elaboration – case example:

An LEA is planning to implement and deploy Artificial Intelligence capabilities to support their analysis of online data during open source intelligence (OSINT) investigations on violent extremism. To ensure they can be held accountable with respect to their AI use, an external oversight body is sought which is specifically empowered to understanding the context of OSINT and AI use. This body has a range of diverse expertise around online investigations, ethics, law and policing at both a national and local level. This will be reinforced through a direct line of communication and engagement with the LEA in question. The oversight body will therefore play a critical role in ensuring that the LEA is being accountable for their AI use during OSINT investigations, which considers the different areas and people affected by the investigation.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data.

4. TRANSPARENCY

Meaning

Transparency is a principle that involves making available clear, accurate and meaningful information about AI processes and specific deployment pertinent for assessing and enforcing accountability. Importantly, the information should establish the necessity and proportionality of any proposed activity involving the use of AI and highlight foreseeable risks.[39] This represents full and frank disclosure in the interests of promoting public trust and confidence by enabling those directly and indirectly affected, as well as the wider public, to make informed judgments and accurate risk assessments.

Practical Considerations

- Who needs to offer transparency? And to whom?
- Maximising transparency should be considered in respect of all stages of the use of AI, from system development to results.
- Account should be made of the importance of the size, nature and source of the datasets being used and the criteria for algorithmic processes, in particular.
- Are public concerns being addressed when making decisions about transparency? Include specific considerations in respect of different sections of society.
- Are there any legal or sector-specific restrictions to achieving transparency? Identify these, along with proposed methods of achieving the aims of transparency in an alternative way.
- Ensure transparency is achieved in a timely, meaningful and appropriate way.
- What processes and criteria are used to judge whether the principle of Transparency has been sufficient complied with?

Of Note

Transparency is fundamental to achieving accountability and the default position should be full transparency or appropriate alternatives that achieve the same aim, in cases where legal or sector-specific constraints apply or in relation to the use of Blackbox AI tools, which are inherently opaque.

Elaboration - Examples for supporting mechanisms:

The mechanism to implementing Transparency is similar to Regulation (EU) No 543/2013 on the submission and publication of data in electricity markets and ENTSO-E central transparency platform (ENTSO-E Transparency Platform). The application and usage of AI by security practitioners by consideration of security measures can be published in a central repository accessible to EU citizens (e.g., EC PCI transparency platform[40]). This can also be implemented at national level as a central repository for all AI usage by the internal security community.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- National industry or sector-specific legal standards relating to public safety.
- National and sector-specific tribunals and formal procedures providing means of effective redress in applicable contexts.

5. INDEPENDENCE

Meaning

Independence refers to the status of competent authorities performing oversight functions in respect of achieving accountability. The oversight body should be independent from individuals and organisations involved in the use of AI including the design, development, supply and deployment. This applies in a personal, political, financial and functional way, with no conflict of interest in any sense. This is an essential condition for effective, credible oversight, as a crucial element in achieving full accountability.

Practical Considerations

- Have effective lines of interaction and communication with the oversight body been established?
- Information being provided to the oversight body must be adequate for the purpose of accountability.
- Determine the nature and extent of independence in a practical sense.
- Determine potential practical or legal limitations to the overall aim of Independence.
- If total Independence is not possible, which form and level of Independence is in/appropriate?
- Does Independence exclude LEAs from accountability bodies?
- How are relationships of the accountability oversight body regulated with pre-existing oversight bodies?
- In case (non-AI specific) accountability processes are already in place, how are relationships between the different accountability processes regulated?
- How will oversight bodies acquire the necessary specialist knowledge to be able to carry out informed, effective decision-making?

Of Note

It may make practical sense to consider existing oversight mechanisms that may be part of the same organisation but operate with guaranteed autonomy. Less than complete autonomy may not necessarily undermine this principle.

Elaboration – case examples:

Police investigating police: "The principle of Independence allows the police to investigate and hold accountable, other activities within the police force. However, the police department undertaking the oversight must not be associated with the department it is investigating. In this instance, you may have a police complaints commission which may have policeman on it. This is typical of the way complaints are handled in the UK, and it is common for one police force to investigate what happens in another force, achieving separation independence." (Source: AP4AI expert input).

New York City: "In New York City, the Algorithm Management and Policy Officer is the executive body empowered for oversight, including for 'receiving, investigating, and addressing any complaints from individuals' about the use of algorithmic systems by public agencies. The Algorithm Management and Policy Officer is a 'centralised resource for agencies, helping provide information about the development, responsible use, and assessment of such tools for the purpose of addressing the risk of inadvertent harm that can accompany them." (Source: NYC AMPO, 2021; https://www1.nyc.gov/site/ampo/index.page)

Examples of Applicable Laws

- National and European laws establishing statutory oversight roles and bodies

6. COMMITMENT TO ROBUST EVIDENCE

Meaning

Evidence in this sense refers to documented records or other proof of compliance measures in respect of legal and other formal obligations pertaining to the use of AI in an internal security context. This principle demonstrates as well as facilitates accountability by way of requiring detailed, accurate and up to date record-keeping in respect of all aspects of AI use. The quality of evidence in this context should mirror that applied to prosecution evidence in terms of integrity, credibility and continuity.

Practical Considerations

- Are processes and procedures in place to allow the capture of the evidence in the required way?
- Are these processes and procedures documented and understood by those who need to know?
- For what purposes might the evidence be used?
- Is it sufficiently robust for these purposes?
- How to define and assess 'robustness', and who is responsible to determine 'robustness'?
- Is the evidence stored in a meaningful and accessible way?
- Is the evidence in its original form subject to legal or sector-specific constraints?
- If so, how can this be managed in respect of achieving accountability and compliance?
- Is the evidence compliant with legal requirements and other principles in respect of being easily understood? If not, how can this be achieved?

Of Note

Depending upon the nature of the evidence, its capture and storage may engage legal and professional restrictions and create the need for appropriate security measures.

Elaboration – case example:

During the use of facial recognition, speech analysis or image analysis algorithms, an LEA recognises the potential for an investigation being heard in court and therefore begins documenting all evidence to form the chain of custody. All areas of the AI system are documented to show how the system recommended a particular decision, alerts of a course of action or proportionate use of these technologies, the response of the investigator and the pictures taken or the speech that was analysed. This information is evidenced and stored following all national evidential procedures. Following the conclusion of this investigation, the case is heard at trial, and the evidence is presented to show how an individual was identified using facial recognition or speech analysis tools. The documentation of the evidence was robust enough, so that it was able to withstand any scrutiny by sufficiently explaining the decisions taken.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- National industry or sector-specific legal standards relating to public safety.
- National and sector-specific tribunals and formal procedures providing means of effective redress in applicable contexts.

7. ENFORCEABILITY AND REDRESS

Meaning

The principle of enforceability requires mechanisms to be established that facilitate independent and effective oversight in respect of the use of AI in the internal security community (e.g., criminal justice context). A crucial aspect of this is to give effect to individuals' fundamental right of effective remedy,[41] established at European Treaty level. It requires that relevant oversight bodies and enforcement authorities have the necessary power and means to respond appropriately to instances of non-compliance with applicable obligations by those deploying AI in a criminal justice context.

Practical Considerations

- Which obligations require enforcement? Distinguish between legal and non-legal obligations in relation to enforcement
- Who determines whether obligations have been fulfilled?
- Which forms of redress will be chosen and how are they related to existing (national, international) redress possibilities? Identify the range of sanctions and remedies, clarifying the conditions relevant to each
- Who determines the appropriate level of redress?
- How is the effectiveness of the remedy determined? Whose responsibility is this?
- Have steps been taken to ensure that the enforceability mechanisms are clearly understood?
- Has the jurisdiction of each organisation been clarified? Is this clearly understood?
- Is information relating to obtaining an effective remedy clear, easily understood and accessible?
- Should those enforcing and implementing redress be independent from each other?

Of Note

Compliance with existing legal obligations, such as those indicated below, is not affected in any way by this principle. In respect of research and development activities, it may be prudent to draft an informal agreement between the relevant parties, setting out duties and obligations in a specified context, including how they will be enforced.

Elaboration

Enforceability and Redress is supported by the principle of Legality, in that laws provide a vital mechanism for enforcement and regulate (some forms of) redress. Also, adherence to Legality will minimise situations for which redress is required. In the same regard, Legality is supported by Enforceability and Redress in that unlawful AI deployments by LEAs are assured to have consequences. The same is true for other AP4AI Principles such as Conduct (where redress can be sought in case of misconduct). Enforceability and Redress is related to Compellability as both ensure forceful mechanisms are in place to guarantee legitimate, safe and

acceptable AI use by LEAs. However, where Enforceability and Redress is focused on enforcing adherence to rules and consequences, Compellability is focused on assuring the access to information.[42]

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- National industry or sector-specific legal standards relating to public safety.
- National and sector-specific tribunals and formal procedures providing means of effective redress in applicable contexts.

8. COMPELLABILITY

Meaning

This principle refers to the need for competent authorities and oversight bodies to compel those deploying or utilising AI in the internal security community to provide access to necessary information, systems or individuals by creating formal obligations in this regard. These specific obligations contribute to the accountability process by regulating the timely provision of relevant, up to date and accurate information in an intelligible format.

Practical Considerations

- The oversight body's role and authority should be determined, which will reflect the degree of access to information that is required in order for them to fulfil their purpose.
- On what grounds can oversight bodies interrupt, interrogate or compel LEAs?
- What mechanisms are used to inform LEAs of and conduct actions related to Compellability?
- What process is in place to clarify and explain what is required, in respect of information and access?
- Have legal and sector-specific obligations in respect of information security been complied with? In what ways has this been achieved?
- What security measures and other safeguards are in place in respect of the provision of information?
- Have the sanctions or consequences of non-compliance been clearly communicated? What conditions apply? How has this been determined?

Of Note

Any restrictions to compliance with this principle should be specific, justified and explained in a clear and meaningful way, as well as forming part of record-keeping.

Elaboration – example of supporting mechanisms:

Implementation can be done through either a new oversight body or by extending the remit of existing bodies such as the European Data Protection Supervisor (EDPS), whereby the EDPS can ensure the safeguarding and the requirements for Compellability. [43]

Examples of Applicable Laws

- National and European laws establishing statutory oversight roles and bodies
- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.

- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.

9. EXPLAINABILITY

Meaning

Explainability is fundamental to the accountable use of AI in a criminal justice context, not solely in terms of the use made of any relevant data sets and processes before a court or tribunal, but also more generally in ensuring that the citizen and their representatives are able to understand, participate and challenge the use of AI. It requires those using AI in this context to ensure that information about this use is provided in a meaningful way that is accessible and easily understood by the relevant participants/audience.

Practical Considerations

- For which aspect(s) of AI or AI usage is Explainability relevant in a specific case? Is there sufficient justification if aspects are excluded from falling under this principle?
- Are clear communication strategies in place that account for different needs of individuals and groups, in respect of the nature and type of information provision? Are processes in place to ensure effective implementation of such strategies?
- Is there clear understanding of the significant risks and consequences of not complying with this principle, either by not providing information or doing so in an ineffective way? How have these been accounted for and mitigated?
- How is the effectiveness of this principle measured? How has this been determined? What factors have been taken into account?
- How to determine whether Explainability has been satisfied? Who judges whether Explainability has been satisfied?
- Have mechanisms to facilitate review, challenge and complaint been established? How is information about these processes been communicated?

Of Note:

The diversity of relevant stakeholders in the accountability process can result in considerable variations in AI expertise or clearance levels. This means that explanations may need to be tailored towards stakeholder groups, while still ensuring sufficient information to make informed decisions. There is further a tendency to value AI expertise before other aspects. However, other forms of expertise such as social or cultural expertise or personal experience with AI impacts are equally relevant to ensure AI accountability can be vouchsafed and thus need to be taken equally seriously.

Examples of Applicable Laws

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.

- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.

10. CONSTRUCTIVENESS

Meaning

This principle embraces the idea of participating in a constructive dialogue with relevant stakeholders involved in the use of AI and other interested parties, by engaging with and responding positively to various inputs. This may include considering different perspectives, discussing challenges and recognising that certain types of disagreements can lead to beneficial solutions for those involved. Being accountable in this way may contribute to building a foundation of trust and confidence in the use of AI, on the part of the public.

Practical Considerations

- What are mechanisms to safeguard Constructiveness in discussions and negotiations?
- Who are the stakeholders that need to be involved?
- How to handle actors that fail to adhere to a basic foundation of Constructiveness?

Of Note

It may be useful to pre-emptively document how particular issues will be dealt with, for example, who is accountable for fixing critical flaws in the AI system should they occur. Security practitioners and oversight bodies should have mechanisms and resources in place to ensure a constructive outcome is given in a reasonable time period.

Elaboration:

Constructiveness should not be misunderstood as non-confrontation at any price ("constructive ambiguity"), as this may cause the stifling of innovation and learning and dilute appropriate challenge. Constructiveness should thus not be confused with 'constructionism' (an established concept in AI). Constructiveness means approaching the relevant aspect of accountability from the perspective of building up rather than pulling down. In this sense Constructiveness can be seen in the reports of auditors and regulators and the language in which their findings are set out. That does not mean there is any dilution of appropriate criticism, sanction or remedy; rather Constructiveness helps shape the focus of the LEAs and their governance bodies when identifying, learning, implementing and reviewing the lessons from experience.

Examples of Applicable Laws and Rules

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.

- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.

11. CONDUCT

Meaning

This principle sits alongside a number of the other principles governing how individuals and organisations will conduct themselves in undertaking their respective tasks and relates to sector-specific principles, professional standards and expected behaviours relating to conduct within a role, which incorporate integrity and ethical considerations. As with the principle of legality this principle extends the formal existing responsibilities in these respects to apply specifically within an AI context, where adherence to these standards is of crucial importance to trust and confidence. The importance of this principle can be seen in one of the key formal instruments to which it relates. The European Code of Police Ethics states, "*the condition of a democracy can often be determined just by examining the conduct of its police.*"[44] 'The conduct of its police' will increasingly include their use of AI technology which therefore represents a specific risk to traditional models of policing by consent – if people withdraw their support for one, they withdraw their support for the other. Where partners in the universal ecosystem are from jurisdictions with different forms of state rule and/or have different values from those of the LEA, there may be a requirement for closer scrutiny and review mechanisms and even barriers to entry into AI programmes involving accountable policing organisations. The expectations of individuals or organisations involved in the relevant AI programme must be expressly identified in advance along with the relevant means that will be used to hold them to account and, in this respect, may vary according to sector, ranging from internal disciplinary proceedings to formal professional sanctions and even proceedings before courts or tribunals.

Practical Considerations

- Are all those involved in the design and deployment of AI aware of obligations in relation to their expected conduct and that of their teams?
- Are existing accountability frameworks in respect of conduct relevant in the context of AI? If not, what modifications need to be made? How will this be determined?
- How are rights and mechanisms of oversight bodies linked to existing LEA-internal processes with respect to AI principles?
- How are complaints documented? What remedies are available to and readily accessible by the complainant? Who is able to bring about a complaint of conduct and to what effect/impact?
- The processes facilitating accountability in respect of conduct should form part of the information made available in respect of specific deployments of AI under the transparency principle.

Of Note

A challenge can be disparities in perspectives of appropriate AI conduct. It is of the utmost importance to clearly identify the ways in which established standards of professional conduct will apply in a specific AI context and/or whether new standards need to be developed.

Elaboration – case example:

A parent reads in an online article about police surveillance that the local police started using an AI-driven pixelization process to blur the images of children whose are in the background when officers use their body-worn cameras to film incidents near schools. The parent thinks that the police should not be using body-worn devices near schools at all and is also concerned that the automated pixelization process is not comprehensive enough to provide any filmed children with the level of privacy that parents and carers would expect. To ensure that the principles of accountability are upheld, the LEA records the complaint, explains how they will look into it and brings in an external oversight body to investigate, not only whether any legal and data protection matters were involved, but also to help her understand the relevant technical and ethical issues involved in her complaint and interpret the response from the police. The individual responsible for investigating the complaint is highly experienced in complaints matters generally but accesses the help of a technical expert within the LEA who agrees the technical parameters of the investigation with the complainant and the oversight body. To uphold public confidence in the matter, the governance body for the LEA is also informed of the investigation and all parties keep in regular contact with the complainant, updating them on all available aspects of the investigation and detailing the process. This maintains an open line of communication should any further issues need to be raised. Following the conclusion of the investigation, the police provide detailed evidence to the complainant and the oversight body who in turn produces a report about the appropriateness of their practices against the Accountability Principles, whether any further policy or guidance is needed and how they reached their decision. This is also communicated publicly through specific, formalised channels in accordance with all the AP4AI Principles to ensure the public are aware of the accountability procedures in place for the investigation of complaints.

Examples of Applicable Laws and Rules

- National, European and International legal instruments, conventions, declarations and agreements specifically pertaining to fundamental rights and freedoms, and secondary provisions relating to identified groups in the same respect.
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.
- National laws protecting or creating individual rights in respect of the exercise of powers by police and law enforcement agencies.
- National criminal justice procedural laws, rules and directions, particularly in respect of fairness, presumption of innocence and the prevention of arbitrary decision-making.
- Professional standards

12. LEARNING ORGANISATION

Meaning

This principle promotes the willingness and ability of organisations and people to improve AI in every respect through the application of (new) knowledge and insights. It applies to people and organisations involved in the design, use and oversight of AI in the internal security domain (security practitioners and partners, industry, oversight bodies, etc.) and includes the modification and improvement of systems, structures, practices, processes, knowledge and resources, as well as the development of professional doctrine and agreed standards.

Practical Considerations

- How do security practitioners learn about/are informed about aspects that need to be adapted?
- How is learning codified to ensure it remains available, replicable and can spread within the organisation/sector?
- Are sufficient resources in place to enable and sustain the learning?
- How will be evaluated whether learning has taken place, goes in the right direction and is sustained?
- Is learning only needed for security practitioners or are other groups equally required to make adjustments?

Of Note

Learning can be challenging to embed into organisations long-term unless some form of codification or structural/cultural embedding takes place and sufficient resources are in place. The establishment of feedback mechanisms is recommended to collect insights such as regular evaluations of current AI practices and of effects of changes to AI deployments. Learning can further be supported by the creation of a 'community of practice' to share AI knowledge and AI practices and the role of established professional colleges, associations and forums is central to the efficacy of this principle.

Elaboration – case example:

An LEA implements AI to help support the horizon scanning and detection of future potential incidents or crime areas to enforce a more proactive community policing approach. This requires a nuanced approach, and the LEA and oversight body acknowledges that there will be particular (localised) challenges or lessons learnt which need to be appropriately addressed. As the LEA is required to make effective decisions in a complex and changing setting, a fast and efficient learning process is needed. Therefore, a learning structure is adopted which runs throughout the whole AI deployment. During deployment, an investigator comes across a recommendation/decision made by the AI system which has potentially discriminatory implications if pursued by the community engagement officers. As a result, the investigator isolates the example and uses it to train existing and new investigators to support them in taking a critical approach engage with the AI system. Not only does this raise awareness across the LEA around the potential challenges that may arise in using AI, but also ensures that LEAs are

being held accountable in their use of AI and how its recommendations/decisions are acted upon. Through the learning approach, this allows the LEA to adapt and modify their behaviour and ensure the correct actions are initiated based on the suggestions raised by the AI.

Examples of Applicable Laws and Rules

- Sector-specific, or organisational established procedures in respect of information security
- National and European legal instruments, conventions and agreements relating to the processing of personal data for criminal justice purposes.

HOLISTIC ILLUSTRATION OF THE POSSIBLE APPLICATION OF AP4AI PRINCIPLES

This section provides a (hypothetical and none-attributed) holistic illustration to elaborate the possible application of the 12 Accountability Principles as an interconnected set of principles as part of the AP4AI Framework. It is envisaged that in the next iteration of this report additional sets of scenarios will be provided for disparate application areas alongside the AP4AI toolkit and implementation guidelines.

A police force is planning to use Retrospective Facial Recognition Technology (RFRT) in order to identify and locate previously unidentified suspects from CCTV camera images over the past 5 years. The RFRT will not be deployed in live public settings but will be used solely to scan the thousands of images retained by the police and compare them to a library of reference images of people who have been arrested. Some local citizens are very concerned about the proposals and the police are already being challenged for buying AI 'spyware' from a company that is associated with human rights abuses and of unjustified/disproportionate interference with the privacy of the citizen via local news and social media. The governance body for the police force is made up of elected representatives and they must assure themselves and those they represent of all the accountability issues before approving the funding for the procurement. Applying the AP4AI principles they are able to determine:

1. **Legality**: That the police have undertaken a comprehensive assessment against all legal and ethical requirements arising from the AI project including those of privacy and data processing (to include all protocols and policies for retention, sharing, deletion, protection and automated decision-making, equality and diversity, human rights, public procurement and contract management and intellectual property. This assessment must include a review of the lawfulness of police's retention of the library of images against which they propose to compare the CCTV

footage and also of the current state of the relevant legislation, guidance and professional practice around facial recognition and surveillance; that all matters of intellectual property, goodwill and commercialisation have been identified and addressed; that they, the governance body themselves, have access to independent and competent legal advice.

2. **Universality**: That all considerations arising from the Principles have demonstrably been applied to the police and all officers and staff for the full life cycle of the AI project and also in relation to all partners including the company providing the technology and/or otherwise taking part in the AI project.
3. **Pluralism**: That the oversight of the AI project is to be achieved through a combination of their own democratic processes (consultation, explanation, revision, challenge), expert technical bodies and advice and informed by the product of forums at local and national/international levels.
4. **Transparency**: That all information necessary to understand and evaluate the AI project and to assess compliance with all of the other principles is, and will continue to be, readily available, published in an accessible form and in a timely way (with any necessary redactions for the purposes of confidentiality/security being given additional scrutiny by identified members having the appropriate expertise and assistance); that meaningful and intelligible records of all relevant decisions and meetings (including accompanying documents considered) are published in an accessible form and in a timely way; that the measures and times by which the project is to be evaluated and audited are clearly published along with milestones and the opportunity to contribute to any review; that there is a clear communications strategy with an identified and accessible contact for seeking further information.
5. **Independence**: That in their conducting accountability and audit processes and activities, they are functionally independent from those whose actions are being held to account and that they are not indirectly dependent on the police, for example, for accessing and interpreting data and resourcing/discharging their accountability obligations.
6. **Commitment to Robust Evidence**: That the AI project is informed by, and conditional upon, robust evidence including the latest professional and academic research literature on AI use generally and on RFRT in particular; that they have a rigorous, independent evaluation of the proposal and that the police will produce and publish the results of a rigorous evaluation to inform the public, sharing existing analytical data which will guide their decision-making and shape their evaluation(s).
7. **Enforceability**: That it has adequate mechanisms by which it will enforce all its obligations within the AI project with clear and accessible processes for remedy and redress.
8. **Compellability**: That it has the ability to require disclosure of all relevant information it deems necessary to undertake its accountability functions in respect of the AI project including access to individuals, products, records and data.
9. **Explainability**: That its members understand the key elements of how the technology works, the risks and benefits of the proposal and that they can explain it to their constituents, separating unsubstantiated anecdotal information from robust evidence; that they are confident in interpreting

and challenging effectively and that the information provided by the police has been understood by all relevant stakeholder groups for whom it was intended.

10. **Constructiveness**: That its deliberations, reports and communications are focused on accountability, emphasising and supporting the other principles transparently and independently, using robust evidence and a commitment to learning and improvement.
11. **Conduct**: That express standards of conduct expected of all individuals and organisations involved in the AI project have been set and will be measured within the project itself, including the ethical considerations arising from suppliers, manufacturers, designers and delivery partners, addressing the citizens' concerns about the company's association with human rights abuses and the perceived development of 'spyware'.
12. **Learning Organisation**: That the police, partners and the governance body themselves are prepared to create, share and transfer knowledge from the AI project internally and publicly (within agreed parameters) and undertake to review and modify their practices, policies, processes and conduct to reflect any 'new knowledge' and insight arising from the project.

NEXT STEPS IN THE EVOLUTION OF AP4AI

This report represents a first important milestone for the AP4AI Project. It defines the initial set of 12 AP4AI Principles based on wide-ranging expert consultations with disparate stakeholder groups across 28 countries.

The 12 Principles serve as a foundation for the upcoming activities towards the realisation of AP4AI's vision. In the upcoming AP4AI activities, the project will provide:

1. A comprehensive review and critical reflection of the existing guidelines, national frameworks, policies and related documentations related to accountability and AI for the internal security community and ecosystem
2. Report on the result of the citizens consultation. The consultation with citizen is a core milestone in the AP4AI Project. The consultation will take place across 27 EU Member States, UK, USA and Australia (addressing approximately 6,000 adult citizens; see section on *methodology*). In this consultation, we ask citizens to review the 12 Accountability Principles, indicate unclear and missing Principles, as well as provide information on acceptable accountability mechanisms. The citizen consultation will allow AP4AI to validate and refine results from the first round of expert consultations, which will provide important reflections and additions moving into the formulation of the AP4AI Framework.
3. Alignment of results from citizen consultation with the AP4AI Framework to ensure a robust and socially acceptable set of principles can be formulated
4. Further scrutiny of principles through continuing expert validations
5. Development of a comprehensive implementation toolkit as an applied mechanism, as well as guidelines for the implementation of the AP4AI Framework
6. Trainings and policy briefings for the internal security community
7. Dissemination of project results and engagement with EU-funded projects, including ongoing and future research projects on AI

APPENDIX A: SUMMARY OF ADDITIONAL EXPERT INSIGHTS

Below we summarise core observations and recommendations collected across expert inputs,[45] including disparities in perspectives which emerged during discussions. These recommendations will inform next steps in the AP4AI Project (see section *Upcoming activities*).

OBSERVATIONS ON THE AP4AI FRAMEWORK GENERALLY

Experts across all stakeholder groups agreed that an Accountability Framework for AI in the context of internal security is a relevant, timely and important instrument, and that accountability as the guiding norm for the security community is appropriate and effective. Moreover, the bottom-up approach chosen in AP4AI was seen as a useful way to capture, understand and subsequently phrase the specific requirements of internal security practitioners in demonstrating accountability, and for capturing the multitude of legal, ethics, technical and citizen requirements.

For the overall ambition they recommended an international Framework with a small number of broad principles to provide the "common ground" from which to build an AI Accountability mechanism. Broad principles will also ensure that the Framework is "future proof", i.e., remains relevant and applicable despite ongoing developments in the AI domain.

To increase usability, experts suggested the use of concrete cases and examples, as well as the creation of practical guidance on processes and mechanisms on how to implement the Framework. Further, experts emphasised the connection of AP4AI work with existing discussions and frameworks on AI (with or without security focus, e.g., EU AI Whitepaper, AI HLEG Ethics Guidelines for Trustworthy AI) that the Framework draws and expands upon, including national approaches.

How this will be taken forward in AP4AI: We are currently conducting a second review of existing frameworks and discussions on AI (cp. section on *methodology*), now that the initial 12 Principles have been established. This review will describe in

more detail the context in which the AP4AI Framework operates and how the AP4AI Framework relates to and differs from other approaches. This analysis will be added in next iterations of this report. For this, we will further engage broadly with experts across all 30 countries for their guidance and advice to ensure that approaches and documents from a wide range of groups and countries are considered. Examples, cases and guidelines are being collected by project partners as part of Cycle 1. Going forward further input will also be sought from our subject-matter experts to ensure examples and cases are easy to understand, correct, unbiased and relevant in the context of AI deployments in the internal security domain.

ROLE OF FUNDAMENTAL RIGHTS AND NATIONAL LAWS

Experts named a number of specific laws that the AP4AI Framework has to adhere to, key among them Fundamental Human Rights and GDPR. Experts were further clear that the AP4AI Framework has to be broad enough to work "with all national legal frameworks." This aligns with our ambition of the Framework as universal mechanism that is applicable across the range of AI usage contexts and situations. Some disparities emerged about the significance of ethics. At the one end were perspectives that suggested a primacy of ethics and explainability "rather than" accountability. At the other end were suggestions that ethics may be insufficient as the basis for an AI Framework, as "ethics is not universal" but can change with time or context. Experts generally agreed however on the primacy of human rights and European AI legislations.

How this will be taken forward in AP4AI: Legal imbedding is a crucial concern for AP4AI, ensuring that its work is clearly in line with EU values and fundamental rights. We are conducting ongoing consultations with experts in human rights, police law and ethics and will continue to validate the Framework with legal experts at every stage. The latter include observation and monitoring of the Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.

MECHANISMS TO ASSURE ACCOUNTABILITY

Experts agreed with Accountability as guiding norm for an AI Framework in the internal security domain. The majority of discussions about accountability suggested a wide range of concrete mechanisms that can help to ensure accountability. Concrete examples of experts' recommendations are given below:

- Accountability needs to be ingrained into the AI system as a means to enforce compliance; designers and developers have a co-responsibility
- Create a certification system for algorithms; ensure auditability of algorithms, data, design processes; offer comprehensive risk assessment frameworks; set a benchmark and/or develop key indicators and quality assurance mechanisms for AI systems to follow security-by-design, privacy-by design principles
- Evaluation by internal and external auditors

- Give warnings to users of potential financial/moral damage; stop AI immediately when causing damage
- LEA should be required to log and trace the collection and use of personal information; user action logging; link to named users
- Usage only after training and signed acceptance of proper usage policy
- Proof of purpose for each AI application
- Develop algorithmic literacy strategies for informing, educating stakeholders; the public needs to know about their rights and that AI is used
- Victim should be provided with effective remedy before national authority

How this will be taken forward in AP4AI: The collection of detailed mechanisms proposed by experts are crucial starting points for the AP4AI toolbox, which will be developed later this year.

CLARIFICATION OF POSSIBLE EXCEPTIONS

Law enforcement and technical experts highlighted that accountability and specific principles (such as Transparency and Compellability) will have to allow for exceptions, in order to satisfy the legitimate and proportionate use of AI by security practitioners to protect citizens. These exceptions relate to AI tools and methods that are classified or situations in which public disclosure of information can harm individuals or investigations. Limitations can also occur when information is located at third parties or in other countries. In the same regard, experts cautioned that limitations and arguments of national security or Intellectual Property Rights may at times be "overstated". Hence, exceptions need to be specific and require clear justification and monitoring by appropriate oversight bodies. Individual experts also emphasised that the scope should be broad, optimally including procurement and trial phases of AI systems.

How this will be taken forward in AP4AI: The detailed application of the Principles at the national level is beyond the scope of the AP4AI project. Yet, we will be able to recommend concrete implementation mechanisms (such as mentioned in the section on Accountability as guiding norm) based on past and future consultations with all stakeholder groups. The core ambition of the AP4AI consortium is to enable all relevant stakeholders (LEAs, oversight bodies, citizens, etc.) to apply the AP4AI Framework effectively and efficiently. Hence, going forward an important milestone will be the creation and implementation of a toolkit and a legal instrument to guide AP4AI applications.

GROUPS RELEVANT FOR AI ACCOUNTABILITY IN THE INTERNAL SECURITY DOMAIN

Experts named a large number of groups and organisations that should play a role in AI accountability processes in the security domain. Table A1 summarises the entries and also highlights relevant disparities in perspectives. The overview illustrates that AI Accountability in the security domain requires a broad approach to accountability, as well as negotiations about how specific groups should participate and at what points in the process.

Table A1. Groups experts mentioned as responsible for accountability oversight

Group	Disparities in expert opinions
Law enforcement agencies (LEAs)	Can be within LEAs; cannot be within LEAs; both inside and outside LEAs
Police ombudsman	
Protection officer/human rights officer	
Interdependency between oversight agencies in policing context with respect to resourcing	
Judicial system	Need more liberty than judges
Prosecutor's office	Should not be prosecutors
People training judges and prosecutors	
Industry	
European Centre, independent technical body	
Public	Although not always directly, can also be through another body; yet not always able or knowledgeable enough
Civil society organisations	
Universities	
Mix/chain of organisations	
Joint taskforce through ministries	
Multiple groups depending on the type of activity in question	
All public bodies must work accountably, not one overarching body; holistic oversight	
Discussion not possible with criminals, suspects or future trespassers	

How this will be taken forward in AP4AI: The extensive list of groups validates the AP4AI principle of Pluralism, as well as the multi-stakeholder approach taken in the AP4AI Project. Going forward, AP4AI will thus continue its close engagement with the broad range of expert groups (cp. section on methodology).

CONTACT

Accountability Principles for Artificial Intelligence (AP4AI)
Website: www.ap4ai.eu
Twitter: @AP4AI_project
Email: Innovation-Lab@europol.europa.eu; CENTRIC@shu.ac.uk

ENDNOTES

1 The European Code of Police Ethics, Recommendation Rec(2001)10

2 Justice by Algorithm - the role of artificial intelligence in policing and criminal justice systems, Council of Europe, Doc. 15156, 01 October 2020.

3 https://www.gov.uk/government/publications/policing-by-consent/definition-of-policing-by-consent

4 Whilst AI is a broad term which has proven difficult to define, for the purpose of this project we have adopted the European Commission High-Level Expert Group definition of AI (2018): "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g., voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g., advanced robots, autonomous cars, drones or Internet of Things applications)." 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 Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final.

5 For example, see proposed EU AI Act, Art 3,17 and 38; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

6 E.g., Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data. Human Behaviour and Emerging Technologies, 1(2), 141-148.

7 See for example PT I of the Police Reform and Social Responsibility Act 2011 in England and Wales.

8 This is in line with Art 38, Laying Down Harmonised Rules on Artificial Intelligence (AI ACT) and Amending Certain Union Legislative Acts; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

9 The initial set will be reviewed in a consultation with citizen across all 27 EU Member States, Australia, USA and UK (currently in preparation under Cycle 2, see section on methodology in this report).

10 For example, see FRA Report 'Getting the future right': https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

11 Cp. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf, p. 9; https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3

12 In the next iteration of this report, a critical review of current frameworks (e.g., 'Ethics Guidelines for Trustworthy AI' developed by the High-Level Expert Group on AI established by the Commission) in relation to AP4AI will be added.

13 Schedler, A. (1999). Conceptualizing Accountability, in: A. Schedler et al. (eds), The Self-restraining State: Power and Accountability in New Democracies (pp. 13-28).

14 Ibid.

15 Thomas Reuters Practical Law. (2021) Accountability Principles retrieved from https://uk.practicallaw.thomsonreuters.com/w-014-8164

16 E.g., Duff, R. A. (2017). Moral and Criminal Responsibility: Answering and Refusing to Answer. Available at SSRN: https://ssrn.com/abstract=3087771 or http://dx.doi.org/10.2139/ssrn.3087771

17 A direct comparison of different concepts compared to AP4AI will be provided in future iterations of his report.

18 Pagallo, U. (2018). Apples, oranges, robots: Four misunderstandings in today's debate on the legal status of AI systems. Philosophical Transactions of the Royal Society A, 376, 20180168.

19 Cp. interview in Computer Weekly, ICO says UK police must 'slow down' use of facial recognition https://www.computerweekly.com/feature/ICO-says-UK-police-must-slow-down-use-of-facial-recognition, 2019; The Guardian, Met's 'gang matrix' breached data laws, investigation finds, https://www.theguardian.com/uk-news/2018/nov/16/met-police-gang-matrix-breached-data-laws-investigation-finds, 2018

20 https://cset.georgetown.edu/publication/trends-in-ai-research-for-the-visual-surveillance-of-populations/

21 https://www.hrw.org/world-report/2022/autocrats-on-defensive-can-democrats-rise-to-occasion

22 The decision of the Court of Appeal for England & Wales on 11 August 2020 serves to underscore the importance of this project. In R (on the application of Bridges) v Chief Constable of South Wales Police and Ors [2020] EWCA Civ 1058 the court identified the key legal risks and attendant community/citizen considerations in the police use of Automated Facial Recognition (AFR) technology during December 2017 and March 2018 and whether those deployments constituted a proportionate interference with Convention rights within Article 8(2) ECHR. The judgment emphasises the critical importance of LEAs having an "appropriate policy document" in place in order to be able to demonstrate lawful and fair processing of personal AFR data. Further, it emphasised that having "a sufficient legal framework" for the use of the AI system includes a legal basis that must be 'accessible' to the person concerned, meaning that it must be published and comprehensible, and it must be possible to discover what its provisions are. The measure must also be 'foreseeable' meaning that it must be possible for a person to foresee its consequences for them (R (on the Application of Catt) v Association of Chief Police Officers [2015] UKSC 9. Each of these elements is covered within this project.

23 Cp. UNODC. (2011). Handbook on Police Accountability, Oversight and Integrity. Criminal Justice Handbook Series. Available online: https://www.unodc.org/documents/justice-and-prison-reform/crimeprevention/PoliceAccountability_Oversight_and_Integrity_10-57991_Ebook.pdf

24 Loader, I. (2016). In search of civic policing: Recasting the 'Peelian' principles. Criminal Law and Philosophy, 10, 427-440.

25 https://www.scottishinsight.ac.uk/Programmes/OpenCall201516/PrinciplesofAccountablePolicing.aspx

26 Planned to be published in March 2022

27 E.g., "Laying Down Harmonised Rules on Artificial Intelligence (AI ACT) and Amending Certain Union Legislative Acts; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206" and other relevant EC documentations based on AP4AI inclusion criteria.

28 Content reviews will be published in future iterations of this report.

29 Fyfe, N., Lennon, G., McNeill, J., & Sampson, F. (2020). The Principles for Accountable Policing. Scottish Universities Insight Institute. More details online at: https://www.scottishinsight.ac.uk/Portals/80/SUIIProgrammes/Accountable%20Policing/Principle%20of%20Accountable%20policing.pdf

30 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

31 Smith, G. (2010). Every complaint matters: Human Rights Commissioner's opinion concerning independent and effective determination of complaints against the police. International Journal of Law, Crime and Justice, 38(2), 59-74.

32 Jones, T., & Newburn, T. (2001). Widening Access: Improving Police Relations with Hard to Reach Groups. Police Research Series Paper 138, Home Office.

33 Loader, I. (2016). In search of civic policing: Recasting the 'Peelian' principles. Criminal Law and Philosophy, 10, 427-440. Walker, S., Police Accountability: The Role of Citizen Oversight. Belmont: Wadsworth Professionalism in Policing Series, 2000.

34 In addition, COVID-19 pandemic restrictions hindered the AP4AI Project to meet experts in a face to face manner.

35 It should be noted that post citizen consultation and validation of cycle 1 (see methodology section) the number of principles or their composition may change. In the next iteration of this report, once AP4AI Principles are finalised, also interlinkages between principles will be detailed.

36 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

37 For example, public procurement guidance

38 In the subsequent iteration of this report, the notion of meaningful participation of public affairs as discussed in A/HRC/39/28 - E - A/HRC/39/28 -Desktop (https://undocs.org/A/HRC/39/28) will be elaborated.

39 Consideration for Application of "A Risk based approach" see https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai in line with EC proposal on AI will be taken to account during the next version of this report.

40 https://cinea.ec.europa.eu/connecting-europe-facility/energy-infrastructure-connecting-europe-facility/pci-transparency_en

41 Cp. article 47 of the Charter of Fundamental Rights: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12016P047&from=EN

42 The interconnectivity and relationships between the principles will be reported in the next iteration of this report.

43 See Appendix A for a list of potentially responsible groups suggested by experts.

44 Recommendation Rec (2001)10 adopted by the Committee of Ministers of the Council of Europe on 19 September 2001, p. 18.

45 Please note that the recommendations and observations cited here do not represent the complete set of comments and insights. Some aspects were summarised or omitted at this point for brevity and clarity. The further work in AP4AI will be based on the full set of observations and recommendations.