



AP4AI

Accountability Principles for AI

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

AP4AI Framework Blueprint



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

AP4AI Framework Blueprint

Version 22 February 2022

Coordinated by:

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

Supporting Partners:

- Eurojust
- EUAA (European Union Agency for Asylum)
- CEPOL (European Union Agency for Law Enforcement Training)

Disclaimer: This report presents the blueprint of the AP4AI Framework. It is structured in three parts: first, it provides a narrative review of current discourse on AI frameworks and regulations relevant for AP4AI; secondly, it summarises high-level findings of an ongoing citizen consultation across 30 countries (5,239 participants to date); thirdly, it provides the first iteration of the AP4AI Framework, with a special focus on proposals for application methods. 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. The project received ethics approval by the university ethics board of Sheffield Hallam University, where CENTRIC is located as academic lead of the AP4AI Project.

Copyright: Copyright notices on individual publications/items must be observed. Unless otherwise stated on an individual publication/item, non-commercial reuse is authorised provided that the source is acknowledged, and the original meaning is not distorted.

Authors:

- B. Akhgar, P.S. Bayerl, K. Bailey, R. Dennis, H. Gibson, S. Heyes, A. Lyle, A. Raven, F. Sampson, *CENTRIC*

Contributors:

- V. Skoric, *European Center for Not-For-Profit Law Stichting, the Netherlands*
- A. Mantelero, *Polytechnic University of Turin, Italy*
- S. Toor, *Helena Kennedy Centre for International Justice, UK*
- M. Gercke, *Cybercrime Research Institute, Germany*

Acknowledgement

The AP4AI Project would like to express its appreciation to the large number of experts who provided their time and insights to our project. We also greatly appreciate the citizens who have engaged with the AP4AI Project and provided us with their valuable insights.

FOREWORD

I am pleased to introduce the “Accountability Principles for Artificial Intelligence in the Security Domain” that you are about to read. This report is the result of intensive work by academics and security practitioners, as well as wide-ranging consultations with renowned experts in the field of AI and with citizens across Europe.

There is growing recognition that research and innovation are crucial in the fight against crime and terrorism. This publication reflects Europol’s strategic commitment to be at the forefront of law enforcement innovation. Law enforcement agencies at national level expect Europol to take a proactive approach to emerging technologies, especially when those technologies are likely to have a profound impact across all jurisdictions. This is clearly the case for Artificial Intelligence, or AI.

This report – and the AP4AI Project more generally – helps to move beyond generalised (and sometimes polarised) treatments of AI. It shows that security actors cannot simply embrace or reject AI; rather, they need to adopt a nuanced approach with the necessary accountability that enshrines fundamental rights.

In addition to the value of the AP4AI Project in its own right, this report is an excellent early example of the value of the EU Innovation Hub for Internal Security. This Hub brings together the different perspectives of the EU Justice and Home Affairs Agencies and the professional communities with whom they work across Europe. Together, the Hub members can identify and prioritise technology topics and bring to bear their wide range of professional disciplines and expertise.

I would like to thank all those who contribute to this project, especially those at CENTRIC and our sister agencies Eurojust, the EU Asylum Agency, CEPOL and the EU Fundamental Rights Agency; and of course, my Europol colleagues.

I am confident that the AP4AI Project will offer invaluable practical support to law enforcement, criminal justice and other security practitioners seeking to develop innovative AI solutions, while respecting fundamental rights and being fully accountable to citizens. This report is an important step in this direction, providing a valuable contribution in a rapidly evolving field of research, legislation and policy.

Catherine De Bolle
Executive Director of Europol

PREFACE

The AP4AI Project is guided by an *enabling* philosophy. The fundamental premise that drives AP4AI and the outcomes it produces is that Artificial Intelligence (AI) is a critical and strategic asset for internal security practitioners.^a The core foundation of the AP4AI Project is that of policing by consent whereby the burden of trust as a mutual obligation between police and society is enshrined within the notion of accountability.

Internal security organisations are data and information centric ‘businesses’, and AI is already an essential element in effective and efficient data and information processing and in converging them into actionable intelligence. Therefore, syllogistically, AI must be indispensable to LEAs and other stakeholders within the internal security domain. AI applications offer crucial support in virtually every step towards resource efficiencies and performance gains – for example, optimising the evidence gathering and analysis process in serious and organised crime cases or aiding the discovery of new adversarial trends and malicious patterns of offending.

The challenge for internal security practitioners is how to capitalise on new technological capabilities that derive from AI in response to societal expectation and demands while, at the same time, demonstrating true accountability and compliance, assuaging societal concern at the use of advanced technology such as AI and automated processing.

This report is the result of intensive multidisciplinary research, extensive engagement with experts and wide-ranging citizen consultation. The project has opted on incremental and live reporting of its outcome rather than traditional end of project reporting. This is to ensure that the project results, outcomes and lessons learned become available to the internal security community within the shortest possible timeframe. It will also allow AP4AI to take practical steps to evaluate its outcome based on the feedback received from internal security domain practitioners, experts including citizens and the research community.

The AP4AI Project, led by CENTRIC and Europol, has the ambition to become a globally known kitemark of quality for research, design, development and deployment of accountable AI use in the internal security domain.

Prof. Babak Akhgar
Director of CENTRIC

Grégory Mounier
EU Innovation Hub for
Internal Security Team
Europol

EXECUTIVE SUMMARY

Artificial Intelligence (AI) is a relatively new technology in the context of policing and on the way to becoming a critical asset for the effectiveness and efficiency of the internal security community, including law enforcement and the justice sector. The challenge for internal security practitioners involved in law enforcement and the delivery of justice is to determine how to capitalise on the opportunities offered by AI to improve the way patrol and response officers, 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. The AP4AI Framework is specifically designed for security and justice practitioners including LEAs and offers validated Accountability Principles for AI as a fundamental mechanism to assess and enforce legitimate and acceptable usage of AI. Its objective is to guide human-centred and socially driven current and future AI capabilities for organisations within the security and justice sector.

This report presents the blueprint of the AP4AI Framework. It is structured in three parts: first, it provides a narrative review of current discussions on AI frameworks and regulations relevant for AP4AI; secondly, it summarises findings of an ongoing citizen consultation conducted across 30 countries; thirdly, it provides the blueprint of the AP4AI Framework, with a special focus on proposals for its application.

Existing frameworks, regulations and debates around AI provide an important basis for AP4AI, which builds and expands upon these works for the specific area of AI Accountability in the internal security domain. Reviewing over 133 documents published since 2017, it was observed that there is no single framework which encompasses the principles necessary to achieve accountable use of AI in the internal security domain. This is problematic given the complex and potentially high-risk nature of AI deployments involved in this area, as it leaves a critical gap for applying a multifaceted, integrated approach to assuring accountability within the internal security domain. This delivers the rationale for AP4AI to establish a coherent Framework to support the internal security domain in achieving AI accountability. The report further outlines vital discussions on Fundamental Rights and their necessity for the integral treatment for AI Accountability. AP4AI integrates Fundamental Rights as key consideration into its Blueprint as core part of the implementation in all Principles.

Contextualisation of the Framework is vital for insights into the cultural, social and political values that determine which principles and implementation processes are meaningful for AI Accountability in the internal security domain across operational and national contexts. This step is thus core to achieving AP4AI's ambition of creating practical mechanisms and tools that directly and meaningfully support AI Accountability. AP4AI has conducted a broad expert consultation in Cycle 1, results of which are reported in the [AP4AI Summary Report on Expert Consultations](#).^b Yet, for AP4AI, core expertise is also located with citizens who are or may be directly affected by AI deployments by security practitioners. After all, we are all citizens irrespective of our chosen occupations, specialisms and qualifications. AP4AI thus conducts a citizen consultation across 30 countries to investigate public expectations of AI accountability and the AP4AI Framework more specifically.

The citizen consultation offers important insights for the further work of AP4AI and the perception of AI use by internal security practitioners more generally. So far 5,239 answers have been collected. The first analysis shows that, while concerns do exist about the AI use by police forces, citizens also see great potential in AI use for safeguarding vulnerable groups and society, including the prevention of future crimes (89.7% agreed or strongly agreed that AI should be used for the protection of children and vulnerable groups, 87.1% that AI should be used to detect criminals and criminal organisations and 78.6% that AI is used to predict crimes before they happen).

There seems further a strong appetite for AI Accountability mechanisms: Over 90% of participants expect police to be held accountable for the way they use AI and for the consequences of their AI use. This suggests that citizens expect strong mechanisms as well as reassurance that policing is willing to deploy AI in an appropriate way. However, only a third of participants (31%) considered existing mechanisms as appropriate. 26% see them as too weak, 9% as too restrictive, while a considerable number of participants (34%) indicated that they "don't know" whether current accountability mechanisms are appropriate. The latter suggests that a considerable part of the public may lack sufficient information about existing mechanisms to make an informed judgement. Citizens showed clear preferences for the groups and organisations which should be responsible for the monitoring and the enforcement of corrects and penalties as part of the accountability process. Courts emerged as the preferred body for both areas, followed by the police themselves and government/ministries. Interestingly, only a relatively small proportion of participants called on citizens to be part of the accountability process, either in a direct process or through participation. Especially for enforcement, citizens were only considered by 9-10% of participants. Least relevant was the inclusion of industry. Groups to be explicitly excluded ranged from citizens to industry, police, criminals, governments and politicians. The citizen consultation also indicates a high acceptance for exceptions, mostly in case of time-critical decisions and if information can help criminals to avoid police, which suggests that citizens are generally sensitive to the complexity of AI Accountability in the internal security domain.

A vital observation is the importance citizens gave to a universal Accountability Framework, rated by over 80% as either important or extremely important to ensure accountability. In addition, all 12 AP4AI Principles were considered highly important to ensure AI Accountability. These results validate the relevance of the 12 AP4AI Principles and gives confidence that constitute an agreed and meaningful foundation for an AI Accountability Framework.

From the outset the AP4AI Project aimed at translating the Accountability Principles (as conceptual representation of AI Accountability requirements) into actionable steps and processes in support of internal security practitioners. In this report, therefore, each of the principles has been qualified with a contextualisation for concrete AI deployment within the internal security domain, providing legal and practical consideration, as well as examples. AP4AI advocates for an AI Accountability Agreement (AAA) that specifies formal and implementable processes for the implementation of the Accountability Principles for different applications of AI within the internal security domain. The AAA should address all AP4AI Principles and their realisation in an operational setting for the specific application of AI. To achieve this, the AAA must include, as a baseline, the four components: context, scope, methodology, and accountability governance. Each phase of the AAA should adopt the application of the twelve principles and use them as a milestone to progress to the next stage. This report describes the AAA and its translation in an operational setting. The report further presents practical considerations for the implementation of the 12 AP4AI Principles, including materiality thresholds, examples of applicable laws, Notes on Human Rights and Data Protection Impact assessment, as well as implementation guide.

This report is a 'living document' (rather than a 'final product') in that it reports on ongoing activities and showcases the current status of the AP4AI Project. This approach is chosen deliberately to allow the collection of feedback throughout AP4AI activities and also to provide full transparency on the project's processes, decisions and approach. Future iterations of this report will emerge as AP4AI updates and refines the Framework towards its core ambition, the provision of a hands-on, practical guidance and tools for AI Accountability.

CONTENTS

Foreword	4
Preface	5
Executive Summary	6
Introduction	10
The AP4AI Project: Accountability Principles for Artificial Intelligence	13
Accountability as guideline for AI use by LEAs and the Internal Security Ecosystem	14
AP4AI approach	16
AI frameworks and regulatory landscape with relevance for AP4AI	18
Approaches by different actors	18
Accountability within existing Artificial Intelligence Frameworks	26
AP4AI Principles in existing documents	27
Legal frameworks with relevance for AI and application in AP4AI	37
Liability in the context of Artificial Intelligence	40
AI and Fundamental Rights	40
Citizen consultation	45
Approach	46
Overview of results	47
Reflection on findings for AP4AI	55
AP4AI Framework Blueprint	57
Outline of mechanisms for the practical application of AP4AI	59
AP4AI Accountability Principles	64
Application Scenarios – Use Case Examples	95
Next steps	102
Appendix A: Details on development of AP4AI Principles	103
Endnotes	112
Project coordination	123
Contact	123

INTRODUCTION

Crimes and criminal organisations become progressively sophisticated, “increasing their operational security by hiding their online activity, using more secure communication channels and obfuscating the movement of illicit funds.”¹ Moreover, criminals are at the forefront of employing innovative capabilities, including Artificial Intelligence (AI), “to facilitate and improve their attacks by maximizing opportunities for profit in a shorter time, exploiting new victims, and creating more innovative criminal business models while reducing the chances of being caught.”²

Internal security and justice practitioners have an obligation to respond to such innovations to ensure they retain the ability to safeguard the societies they serve. One of the capabilities that security practitioners are employing for this purpose is Artificial Intelligence.³

AI is a relatively new technology in the context of law enforcement but on the way to becoming a critical asset for the effectiveness and efficiency of the internal security community, including law enforcement and the justice sector.⁴ 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 optimising the evidence gathering and analysis process in serious and organised crime cases or by aiding the discovery of new adversarial trends and malicious patterns. Accordingly, the uptake of AI is growing with a wide variety of different AI techniques across national contexts and law enforcement priorities.⁵

In the same regard, 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

Internal security institutions are entrusted by society with the restriction of personal freedoms for the pursuit of enforcing the law and the provision of safety and security. However, they are able to operate effectively only to the extent that they possess and retain the public’s confidence they have the mandate to protect. The usage of AI by internal security practitioners thus needs regulation

and clear guidance to ensure that the use of algorithms and AI-based systems and platforms are not only carefully understood but also 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 is to determine how to capitalise on the opportunities offered by AI to improve the way patrol or response officers, 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**. The AP4AI Framework is specifically designed for security and justice practitioners including law enforcement agencies and offers validated **Accountability Principles for AI as a fundamental mechanism to assess and enforce legitimate and acceptable usage 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, AP4AI offers a step-change in the application of AI by the internal security community. Its objective is to guide human-centred and socially driven current and future AI capabilities for organisations within the security and justice sector.

AP4AI will deliver concrete products to support internal security practitioners in their deployment of AI:

- A robust set of agreed and validated Accountability Principles for AI, which integrate practitioners as well as citizens' positions on AI8
- Implementation guidelines and toolkit including supporting software tool to give practitioners and oversight bodies concrete, practical, actionable compliance and assessment tools to assess and review AI capabilities from design to deployment
- Trainings and policy briefings for the internal security community and oversight bodies on how to apply the AP4AI Framework, as well as broader insights from AP4AI research
- A set of reports and documentations as reference for the internal security and judiciary community, as well as oversight bodies and the public
- Engagement with national and EU-funded projects to inform ongoing and future research efforts on AI with respect to AI Accountability needs and applications

This current report constitutes the first iteration of the AP4AI Framework and provides a blueprint focused on mechanisms for the implementation of AI accountability. It expands on the previously published report on AP4AI Principles in three ways.⁹ First, it conducts a narrative review of existing documents, frameworks and regulations on AI to clarify the embedding of AP4AI in current AI discussions and its innovations; secondly, it presents high-level findings of the ongoing citizen consultation across 30 countries (the 27 EU Member States, Australia, USA and UK, collecting 5,239 answers so far); thirdly, it outlines first steps towards the practical implementation of the Principles which is at the heart of AP4AI's efforts. The practical guidance will see ongoing developments and refinements, as AP4AI will continue its engagement with all expert groups for validation, expansion and contextualisation (see see section on [section on AP4AI approach](#)). Moreover, AP4AI takes the position that any AI framework needs to be a 'living document' to accommodate the continuously changing landscape of AI developments. Readers can thus expect important additions towards the practical implementation of the AP4AI Framework in future iterations of this report including domain-specific implementation guidelines (e.g., on Child Sexual Exploitation, Counter Terrorism, Serious and Organised Crime and protection of public spaces) as well as the process modelling of the AP4AI Principles.

THE AP4AI PROJECT: ACCOUNTABILITY PRINCIPLES FOR ARTIFICIAL INTELLIGENCE

The AP4AI Project was created in recognition of the complexities, that the law enforcement and justice sector faces in researching, developing, procuring and deploying AI capabilities in societies that, on the one hand, rightly expect to be protected by the best means possible, and on the other hand, justifiably request that these means do not impinge on societal freedoms and individual rights. Achieving this balance is a complex and ongoing challenge, which requires continuous negotiations between disparate expectations and needs.

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 creates a Framework for security and justice practitioners including 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. In this way, the AP4AI Framework will allow practitioners to capitalise on available AI capabilities, whilst demonstrating meaningful accountability towards society and oversight bodies.

AP4AI focuses on accountability as a guiding standard under the premise that in the field of security and justice, functional 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 utilizing AI and (b) oversight bodies and the public to measure security practitioners' use of AI against.¹⁰

The AP4AI Framework will provide a mechanism to proactively assess, as well as reactively demonstrate AI Accountability. 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 applications by LEAs, justice agencies and their partners whichever domestic jurisdiction they operate within.

The conceptual foundation of the AP4AI Framework is a set of 12 Accountability Principles (see [section on AP4AI Framework Blueprint](#)). These Principles were developed in Cycle 1 of the AP4AI Project in collaboration with multi-disciplinary and international subject-matter experts.¹¹ The 12 Accountability Principles are meant to inform legislative bodies to create future-proofing legislation and enforcement directives agnostic of particular technological changes. The AP4AI Framework is introduced in this report and will continue to be refined, validated and applied in different practical setting (i.e., use cases) in subsequent publications.

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

AP4AI 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".¹² Following this concept, AP4AI advocates 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), deployments (e.g., systems, software platforms, usage situations), as well as communities and individuals that are (potentially) affected or involved.

Accountability comprises in itself the three aspects of monitoring, justification and enforcement,¹³ and in a legal perspective is defined as the "acknowledgement and assumption of responsibility for actions, decisions, and their consequences."¹⁴ 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, learning and improvement.

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',¹⁵ 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 the 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.

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 enforceable obligations and thus provides a foundation that has actionable procedures at its core. The notion of accountability therefore offers vital benefits compared to other instruments and frameworks.

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 level¹⁶ 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 broad and sustained engagement.¹⁷

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 APPROACH¹⁸

AP4AI is built on an expert-driven approach that emphasises the broad engagement with actors across society. AP4AI's expert-driven approach ensures that its results are grounded in a comprehensive understanding of Accountability and Artificial Intelligence in the internal security domain and developed based on a comprehensive set of perspectives, expectations and requirements. So far, the project brought together expertise from LEAs and border police, justice and judiciary, human rights, ethics, industry, and civil society across 30 countries.¹⁹ The international setup of the consultation 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. Most importantly, the project consults and engages 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.

In consequence, AP4AI solutions are specifically designed to support the wide range of disparate stakeholders that are taking part in or are affected by AI deployments within the internal security domain, i.e., practitioners in the security, policing and justice domain, oversight bodies, law makers, industry, researchers and research institutions, as well as citizens. AP4AI supports these stakeholders in researching, designing, assessing, reviewing and revising 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 in line with EU values and fundamental rights.

To ensure the robust development and validation of the AP4AI Framework and products, the project is conducted in three cycles as consecutive steps of exploration, integration and validation. These three cycles build on each other to ensure effective integration of perspectives across stakeholder groups:

- **Cycle 1 – Development of the AP4AI Principles (completed):** The first cycle consisted of two activities: (a) a review of over 130 existing frameworks aiming to guide or regulate AI and (b) expert consultations with 69 subject-matter experts from law enforcement, justice, legal, ethical and technical fields identified by the AP4AI partners. Results of the expert consultations are reported in the [AP4AI Summary Report on Expert Consultations](#).²⁰
- **Cycle 2 – Citizen consultation for validation and refinement of the Principles (ongoing):** An online consultation is being conducted in 30 countries (all 27 EU members states, UK, USA and Australia) to collect citizen input on the AI Accountability Principles developed in Cycle 1, as well as insights into possible accountability mechanisms. The intended audience is 6,400 citizens. At this point, the majority of answers (5,239) has been collected and indicative results are presented in this report (see [section on Citizen consultation](#)). Once the consultation is completed, citizen results will be integrated with Cycle 1 results to inform and refine the AP4AI Framework.

- **Cycle 3 – Expert consultation for validation and contextualisation of the AP4AI Framework (upcoming):** The AP4AI Framework will go through continued validations by 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. Activities in Cycle 3 will include structured feedback collection, hands-on implementation workshops, as well as case creation for the operationalisation of the Framework into practice. These consultations will refine and contextualise the AP4AI Framework. Consultations, contextualisation and refinements will be an ongoing and continuing process to ensure that AP4AI solutions reflect new developments and emerging trends in AI deployments. It is envisaged that in this cycle the AP4AI Project develops a software tool to support internal security practitioners in the realisation of the AI Accountability Agreements (see [section on AP4AI Framework Blueprint](#)).

AI FRAMEWORKS AND REGULATORY LANDSCAPE WITH RELEVANCE FOR AP4AI

AI is well-positioned to play a vital role in all sectors of society.²¹ Preparing for the use of AI within society has therefore become an issue generating significant debates and discussions within public and scholarly discourse,²² leading to a large body of work that aims to provide expertise as well as regulatory guidance. These efforts are driven by the recognition of “fierce global competition”,²³ and “significant gap[s in] understanding how to make sense of existing laws, regulations and ethical standards”²⁴ creating repeated calls for overarching frameworks, including those for law enforcement, that also discuss the status of accountability.²⁵ These efforts provide an important basis for AP4AI, which builds and expands upon these works for the specific area of AI accountability in the internal security domain.

APPROACHES BY DIFFERENT ACTORS

A large proportion of existing frameworks are broadly focused and aimed towards the private sector or businesses, instead of the security domain.

Our review of over 130 documents identified only 18% with an explicit focus on this area (cp. [Appendix A](#)). The broad scope of most frameworks has met with concern from experts in relation to their adaptability and transferability into the internal security sector considering the specific challenges of this area.²⁶ However, this is not to say that they cannot provide important insights into accountability which can be re-contextualised by AP4AI to be effective within the area of internal security.

In the following we reflect on core observations from documents across various stakeholder groups (governmental and administrative bodies, law enforcement and internal security actors, industry and technical interest groups, civil society organisations and academia) with relevance for AP4AI.

Governing and administrative bodies

With governing and administrative bodies, we refer to a group of organisations which implement rules and govern the actions and conduct of the public sector. Governing and administrative bodies generally agree on the importance of accountability. In 2017, the UK Information Commissioner's Office (ICO)²⁷ proposed that accountability should become an explicit requirement under the Data Protection Act 2018 and GDPR, meaning in extension that it must be recognised as a legally binding concept also for the internal security domain.²⁸ Accountability is also mentioned as a core ethical principle by bodies such as the Law Council of Australia,²⁹ the UK Government,³⁰ the European Commission³¹ and the European Parliamentary Research Service.³² These publications place a high value on accountability, although frequently as one amongst other principles. Nonetheless, they illustrate the broad appeal of accountability as basis for AI guidance and support the AI Accountability focus of AP4AI.

Considerations by governing and administrative bodies tend to demonstrate an ethically-driven focus in their discussions.³³ This can be seen in numerous national strategies which established ethical regulatory AI frameworks and AI ethics committees and councils.³⁴ The same focus is also visible in various calls for building public trust in the design, development and implementation of AI systems, as well as for calls to situate them within socio-cultural contexts.³⁵ The ethics focus underlies also the concept of 'Trustworthy AI', as formulated by the High-Level Expert Group on AI (AI HLEG).³⁶ The AI HLEG Guideline suggests a process-orientated approach which encompasses technical and non-technical implementation methods based on a three-tier system. Included is a list of seven requirements to achieving Trustworthy AI, which also includes accountability. The authors moreover emphasise the importance of practical application in that "mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use."³⁷ In this sense, the Guideline provides vital pointers on realising and assessing trustworthiness. In the same regard, the document is focused primarily on ethical concerns, i.e., it does not provide legal guides and legislative grounding, as relevant to applications in the internal security domain.

Due to the sensitivity of AI use by policing and law enforcement, ethical, legal and human rights principles throughout the AI cycle are generally given special consideration.^{38,39} For instance, Fuster and colleagues identify ethical guidelines and fundamental rights as "the foundations of Trustworthy AI."⁴⁰ However, the attention given to law has been labelled as vague, leading to a potential blurring of boundaries between ethical and legal parameters.⁴¹ This can be problematic for the law enforcement domain, if it remains unclear whether accountability principles are based on ethical or legal foundations, which may subsequently lead to frameworks being underused. Our own expert consultations further cautioned against a (sole) reliance on ethics for AI frameworks, as ethic values tend to be 'fluid' and dependent on personal values, contexts or historical settings (cp. [AP4AI Summary Report on Expert Consultations](#)).⁴² For a critical, potentially high-risk area such as the internal security domain, foundations for any AI framework should instead be based on a more 'stable' foundation, such as laws and fundamental rights. Judiciary bodies have specifically emphasised the role of fundamental human rights based on values such as non-discrimination and transparency.^{43,44}

Findings from our expert consultations thus largely align with concerns in current discussion, leading to the adoption of Legality as is the foundational principle within the AP4AI Framework (cp. [section on AP4AI Framework Blueprint](#)).

Frameworks across most governing/administrative bodies suggest that fulfilling tenets of 'Trustworthy AI' will have cascading benefits in building public trust and resolving existing "trust hurdles"^{45,46,47,48}. However, they also rightly suggest that due to the 'context specificity' of AI systems sectorial approaches may be needed. This is reflected in a joint report by Europol and Eurojust (2019),⁴⁹ which identified challenges relating to national legal frameworks in international criminal investigations and prosecution of cybercrime. It states that dedicated legislation that specifically regulates law enforcement presence and actions in an online environment, along with forensic-technical standards for the collection and transfer of e-evidence, should be further developed, promoted, and adopted. In addition, the report argues that legislation should be harmonised at EU level, allowing for more effective joint operational actions such as large-scale botnet and/or underground criminal forum takedowns. While not directly AI related, such observations are extremely relevant for the context of AI deployments in the internal security domain, as the broad scope of most current guidelines does not make reference to the specific requirements in this area. Ensuring that frameworks and underlying principles are context-specific and work in harmony is key to implementing a successful framework. AP4AI reacts to this observation by targeting its work towards the internal security and justice domain, developing its framework for and with internal security practitioners and subject matter experts in the domain.

Although approaches by states and governing bodies tend to focus on similar arguments, a key recommendation by the Committee on Standards in Public Life⁵⁰ remains valid, namely that frameworks need to be made clearer and easier to navigate to support their translation into practice. This recommendation echoes the concern that most frameworks are not specific enough to be applicable for the unique nature of law enforcement and criminal justice systems causing challenges for their successful adoption.^{51,52} Furthermore, where frameworks have centred around values such as ethics and wide, generic target audiences such as the public, this can lead to an under-emphasis on the binding instruments which guide law enforcement practice. In consequence, such frameworks have been labelled as too "abstract" or "vague", which can reduce their actionability within a security setting.⁵³ Instead, AI frameworks should combine common principles with tangible solutions which are dynamic and scalable in nature,⁵⁴ to ensure that frameworks are sufficiently flexible to accommodate AI solutions within ethical, legal and societal contexts while still providing clear boundaries.⁵⁵ This is not a straightforward task, as the purpose, focus and proposed outcome of AI use within the internal security domain can differ amongst nation states as well as local and regional contexts.⁵⁶

Taking these recommendations to heart, AP4AI makes a concerted effort to create practical and validated guidance for internal security practitioners as well as an AI Accountability Agreement as an implementation enabler, which will support the realisation of AI Accountability within individual operational contexts (cp. [section on AP4AI Framework Blueprint](#)).

Law enforcement and internal security actors

Discussions in the law enforcement and internal security domain illustrate growing awareness that it is facing a “changing trust landscape”⁵⁷ in deploying technological innovations. Unsurprisingly, law enforcement practitioners have shown a profound interest in approaches which reinforce positive public-law enforcement relations.⁵⁸ This can be seen in a number of documents which highlight the importance of establishing police accountability in general. For example, the UNODC *Handbook on Police Accountability, Oversight and Integrity* identifies transparency, openness to scrutiny, integrity and assuring public confidence and legitimacy as the four core attributes to ensure accountable policing.⁵⁹ Accountability has also been described as a central theme within the *Police Scotland Policing 2026 Strategy*, which discusses strategies to maintain legitimacy and relevance.⁶⁰ Similarly, the *UK Police Foundation* labelled accountability as fundamental in allowing citizens to challenge new policing practices.⁶¹

Yet, although AI and accountability in policing have become a central point of discussion across the law enforcement and internal security sector, they are often discussed in isolation and not as a targeted approach to ensuring accountability for AI deployments. This means that there remains a significant gap in addressing AI Accountability within the fields of security and policing.

This accountability gap has been recognised by the internal security community, which considers the responsible use of AI as one of the central questions of modern policing.⁶² For example, the *National Security Commission on Artificial Intelligence* outlines a process for ‘Responsible AI’ and for the creation of confidence in the adoption and use of AI. The document lists key aspects which must be addressed: (a) robust and reliable AI, (b) human-AI interaction and teaming, (c) testing and evaluation, verification and validation, (d) leadership and (e) accountability and governance. With respect to accountability and governance, they note that government agencies need to adapt existing accountability policies to the AI lifecycle, whilst establishing new policies which allow for concerns about irresponsible AI development and use to be raised.⁶³ This is an important facet to consider, particularly as it concerns the incorporation of auditing and reporting, review mechanisms and appeals and grievance processes.

The concept of risk assessment and review processes are not uncommon within the internal security sector, which has been identified as a key mechanism for ensuring that AI systems can be monitored. This is supported within the literature, which stresses the importance of mandatory testing across the AI development stages to mitigate against risks.⁶⁴ However, proposals tend to lack details in how this will be implemented in practice, specifically considering the various actors involved in the use of AI across law enforcement and the internal security domain. *Moreover, the risks of failing to use available AI-driven solutions in the prevention of serious criminal offending are rarely considered.*

The AP4AI Framework integrates such approaches in the development of a more detailed enforcement of Accountability that considers the complexity of modern law enforcement and internal security actors. More specifically, it emphasises Compellability, as well as Enforceability and Redress to allow scrutiny and to enshrine the right of the public for rectification. AP4AI further puts strong emphasis on understanding and engaging the public, not only as a basis to build trust but to open channels for citizens to have an active voice in the AI Accountability process (see [section on AP4AI approach](#)).

Industry and Technical Interest Groups

Industry and technical interest groups focus strongly on data responsibly, whilst militating against potential harms to end-users by means of upholding ethics, human rights, privacy and security concerns. However, discussions seem to lack common values and norms, as well as the tools and mechanisms to operationally implement these principles, most importantly accountability.^{65,66} In this respect, technical approaches seem arguably less advanced and entrenched, when compared to their public-facing counterparts.

More recent industry efforts have shown positive attempts to addressing these issues. The *Partnership on AI* is a clear example, which examines the intersections between AI with societal-focused principles such as fairness, transparency and accountability.⁶⁷ It aims to answer questions related to core AI issues such as equality, explainability, responsibility and inclusion. The *Partnership on AI* group is exploring these issues through projects like ABOUT ML (Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles).⁶⁸ Such projects show that industrial discussions are fast advancing and will benefit the design and development of AI systems. A gap that remains also in industry discussions is how the principles, and specifically accountability, can be translated and will be maintained once AI is adopted by high-risk sectors such as the internal security community.

The ethical influence is also identifiable in considerations by industry actors. As noted by IBM, “it is imperative to understand the ethical considerations of our work.”⁶⁹ To achieve this, IBM proposes five areas: accountability, value alignment, explainability, fairness and user data rights. Samsung equally names fairness, transparency and accountability,⁷⁰ while Microsoft adds inclusiveness, reliability, safety, privacy and security.⁷¹ Another example is Accenture, which provides the ‘four pillars of Responsible AI’, namely organisational, operational, technical and reputational.⁷² These pillars are discussed through a series of recommendations and case studies to understand how they can be achieved in practice. For instance, the reputational pillar refers to ensuring companies are achieving Responsible AI as linked to company values, ethical parameters and accountability structures. To implement this, Accenture argues that risks are managed through pressure testing and an Algorithmic Assessment toolkit.

However, in contrast to other groups, industry discussions focus primarily on AI designers and developers and industrial self-regulation.⁷³ Also, industrial approaches tend to be more narrowly defined. They tend to cover primarily the beginning of the AI design and development process, often omitting assessment

and review processes as part of meaningful accountability. For example, responsibility is often seen to lay with AI designers and developers, who should keep clear records of company policies, actions, software outreach and business conduct guidelines.⁷⁴ Furthermore, industrial guidelines tend to lack details on how law and Fundamental Rights should be applied in ensuring accountability from a technical perspective. This results in AI approaches being designed with a specific “corner” of AI in mind,⁷⁵ raising questions about their direct applicability across the full AI lifecycle and the internal security domain.

Non-governmental organisations and civil societal actors

The role of NGOs and civil society actors in supporting the regulation of AI development and use has been widely noted across literature.⁷⁶ The Alan Turing Institute, for instance, highlights the importance of incorporating accountability across the AI lifecycle.⁷⁷ They further argue that the “accountability gap” (i.e., AI systems not being morally responsible in the same manner as humans) must be addressed by clearly identifying which individuals should have responsibility within the AI production line. AI Now similarly argues that “public agencies urgently need a framework to assess automated decision systems and to ensure public accountability.”⁷⁸ This emphasises the importance of implementing human accountability through both answerability (by justifying decision-making processes) and auditing (determining who is responsible for a particular component or action).

As noted by the Committee on Standards in Public Life⁷⁹, there are a number of ethically-focused approaches which are widely cited in the use of AI in the public sector.^{80,81} This prominence of ethical approaches is reflective of the overarching concerns which started the debate on AI governance.⁸² One example is the *Ethical Platform for the Responsible Delivery of an AI Project*,⁸³ which focuses primarily on FAST (fairness, accountability, sustainability, transparency) Track Principles and SUM (support, underwrite, motivate) Values which together are meant to ensure an ethically sound and reliable AI system. A highlight of this framework is the establishment of a Process-Based Governance Framework, which users should use to integrate the FAST Track Principles and SUM Values throughout the implementation of AI within an organisation. However, the guidelines largely focus on “big picture issues” and “user centred requirements”, which will not have equal standing across AI projects and systems.⁸⁴ This can cause challenges in understanding how this can be implemented explicitly within the internal security domain, as well as in noting how the principles explain the systems used.

Another example are the AI principles proposed by the OECD as a legal instrument for governments to create a human-centric approach to achieving ‘Trustworthy AI’ as conceptualised by the EC Ethics Guidelines for Trustworthy AI.^{85,86,87} The principles are value based, consisting of (a) inclusive growth, sustainable development and wellbeing, (b) human centred values and fairness, (c) transparency and explainability, (d) robustness, security and safety and (e) accountability.⁸⁸ Combined, these principles do provide a harmonised, ethically driven approach to achieving Trustworthy AI.

A third example is the IEEE Ethically Aligned Design report,⁸⁹ which identifies human rights, the prioritisation of wellbeing, accountability, transparency and awareness of misuse as overarching principles. The report also includes a comprehensive set of concerns and recommendations which fall under the areas of legal status of AI, government use of AI, legal accountability for harms caused by AI, transparency, accountability and verifiability in AI. Only one issue was raised in relation to law enforcement, namely: *“how can AI interact with government authorities to facilitate law enforcement and intelligence collection while respecting rule of law and transparency for users?”*⁹⁰ There is consequently a demand for a framework which is primarily designed at addressing the complex, context-specific issues which law enforcement faces in AI use.

Other frameworks propose principles which have a stronger focus on the actual design, use and implementation of AI in practice. The Center for Democracy and Technology provides a practical infographic which encourages AI practitioners to consider the thought processes behind the design, building, testing and implementation of AI.⁹¹ In another example, the Oxford Commission on AI and Good Governance identifies inclusive design, informed procurement, purposeful implementation and persistent accountability as the four overarching principles to achieving good governance.⁹² This is also evident in the AI Standards Roadmap, whereby concepts such as privacy, inclusion, safety and security-by-design are considered fundamental within Management System Standards to achieve Responsible AI.⁹³

The discussed approaches align with best practices upheld by AP4AI which emphasises multi-level participation (encoded in the Pluralism Principle), mechanisms to support implementation and a comprehensive process view on the AI lifecycle. The latter also emerged as part of the AP4AI expert consultations (e.g., to also cover piloting phases by AI Accountability mechanisms in the same way as operational deployments).⁹⁴ Where such approaches fall short is again in lacking a law enforcement-facing perspective, meaning it remains unclear how the individual principles should or can be achieved within an operational context. This emphasises the key ambition for AP4AI, which places public engagement and practical guidance at the centre of its efforts.

Academic approaches

Academia has made significant contributions to addressing the issues posed by AI. This has been achieved through both the evaluation of existing approaches and by developing new propositions, although the areas covered show considerable similarities with other groups. Reviewing past approaches, Joblin et al.⁹⁵ identified ten recurring ethical values across 84 policy documents: transparency, non-maleficence, responsibility, privacy, beneficence, freedom/autonomy, trust, sustainability, dignity and solidarity. Similar findings emerged in reviews by Hagendorff⁹⁶ and Beckley and Kennedy.⁹⁷

However, scholars have also put forward criticisms to the primary focus of ethically centred approaches. Hagendorff, for instance, argues that ethical principles are broad and overarching in nature, which are then required to be implemented into a diverse set of practices and geographical groups which have different

responsibilities and priorities in AI use. As a result, “ethics thus operates at a maximum distance from the practices it actually seeks to govern.”⁹⁸ This is supported by Mantelero and Esposito⁹⁹ who argue that regularly used ethical principles are problematic as they do not refer to the longstanding legal principles which already have been contextualised within different fields. For example, product safety and data governance are not purely ethical and must be recognised as legally binding requirements. Some scholars have further suggested that existing approaches may be limited in keeping up with technological advancements. Specifically, existing tools available to the internal security domain, where transferred from other contexts, often fail when applied to AI systems.¹⁰⁰

Moving beyond critical reviews of past work, scholars have also proposed new approaches to addressing AI challenges. Doshi-Velez and Kortz,¹⁰¹ for instance, examined how AI systems can be held accountable through the principle of explainability, which is widely covered across the AI literature.^{102,103,104} They argue that ‘Explainable AI’ is important to ensure that AI systems can be scrutinised. This aspect is noted by Coeckelbergh as an epistemic dilemma, whereby approaches have aimed to overcome the knowledge problems posed by AI.¹⁰⁵ This is a justified argument, particularly in the context of the internal security domain with potentially high-risk AI applications and where output of AI systems may be needed as evidence in court.¹⁰⁶ Yet, existing proposals again do not provide practical solutions as to how they can be implemented in such a contextualised setting.

As another example, Schrader and Gosh¹⁰⁷ offer a social and ethical framework which recommends that AI systems must be proactively designed and developed with consideration of factors such as ethical issues, human awareness, collaboration with AI, accountability and AI integrity in mind. Again, these are important points which should be considered. However, the model does not provide any focus within a specific organisation or institution.¹⁰⁸ This is important, as an integrated approach which looks at the entire system and how these systems interact is fundamental to assessing the project and its impacts.¹⁰⁹ As emphasised by Freeman et al., achieving meaningful accountability is rooted in “concrete and technically-informed thinking within and across contexts.”¹¹⁰

Overall, it can be observed that there is no single framework which encompasses the principles necessary to achieve accountable use of AI in security and policing across jurisdictions. This is problematic given the complex and potentially high-risk nature of AI deployments involved in the security and policing domain. This means that there are no concrete guidelines for those researching, developing, operating and/or assessing AI in the security and policing sectors to ensure AI Accountability. The European Parliament has similarly observed that present discussions around an AI regulatory framework have so far only focussed on the Digital Single Market agenda, not considering in detail the singularity of law enforcement and criminal justice, both regarding the specific risks related to their use of AI and the peculiarities of the EU legal framework in the Area of Freedom, Security and Justice (AFSJ).¹¹¹ This lack of an overarching framework has the potential to lead to a fragmentation in AI practices and lacking accountability overall.

ACCOUNTABILITY WITHIN EXISTING ARTIFICIAL INTELLIGENCE FRAMEWORKS

As the review of the manifold approaches and discussions to AI indicates, accountability is a recurring ambition by actors varying from EU Commission and its high-level expert group on AI¹¹² to governing bodies within the UK¹¹³, Australia¹¹⁴, US¹¹⁵ and across the EU¹¹⁶ to industry¹¹⁷, civil society group¹¹⁸ and academic think tanks such as Ada Lovelace Institute¹¹⁹ and the Alan Turing Institute.¹²⁰ Reviewing the definitions of accountability, it becomes apparent that they draw upon repeating elements, most importantly: redress, effective oversight mechanisms, auditability, submission to scrutiny and the right to be made aware of and challenge the implementation of AI systems.

- **Facilitating redress:** Facilitating redress is a core feature of accountability definitions, suggesting that accountability should include an opportunity for those affected by AI systems to rectify or remedy its impacts. The element of redress is evident across all domains, including definitions from NGOs,¹²¹ research institutions¹²² and governing bodies¹²³ but seems to be less prevalent in the security domain.
- **Effective oversight mechanisms:** Closely linked to facilitating redress are effective oversight mechanisms to enable the ruling and implementation of accountability. Effective oversight mechanisms within accountability measures suggest that those conducting oversight should also implement appropriate mechanisms allowing for accountability to be assured. This element is evident across all sectors¹²⁴ including AI frameworks with an accountability component published by security organisations.¹²⁵
- **Auditability:** Auditability with respect to AI practices considers that in both the design and implementation stage, the data and processes which underpin the AI system's decision-making should be reported to the highest standard. This will allow for effective traceability whilst also feeding into the system's transparency. This element was widely adopted in definitions by European governance bodies,¹²⁶ technology groups¹²⁷ and research institutions,¹²⁸ but less prevalent in documents addressing the security domain.
- **Submission to scrutiny:** Submission to scrutiny suggests that those who design and implement AI systems, along with those who undertake the review process, should open themselves up to the public and to oversight bodies for assessment and review. Across definitions, submission to scrutiny (i.e., the inspection or investigation of AI systems and its potential impacts) was raised as an important factor for a range of actors including the public, courts and relevant oversight bodies. Evident within definitions across all sectors,¹²⁹ this element was also an important feature of AI frameworks published by security organisations.¹³⁰
- **The right to be made aware of and challenge:** The right to be made aware of and challenge the design and implementation of AI systems was an equally recurring feature. This element purports that all actors, but specifically those which the AI system may affect, should have all information related to the AI system including all stages of its design, its implementation mechanism and its expected and actual impacts. This element was identifiable within the definitions of security organisations, along with a variety of actors from NGOs¹³¹ and administrative authorities.¹³²

As the above demonstrates, definitions of accountability generally assume similar meanings across sectors. The central tenet put forward across all considerations is the need to hold individuals and organisations accountable for their AI usage and the consequences of their AI systems.

A recent description of accountability specific to the internal security domain similarly refers to holding individuals and organisations responsible for the design, oversight and implementation of AI systems with all actors (including the public and external organisations) having the right to know and understand the system, its oversight and impacts.¹³³ Yet, compared to other sectors, it does not draw on further aspects such as auditing, multi-level participation, and the facilitation of redress and certification.

For AP4AI, this leaves a critical gap for applying a multifaceted, integrated approach to assuring accountability within the internal security domain. Also, current approaches, while proposing accountability as an important requirement, tend to fail in providing the conceptual and practical tools to its implementation. While accountability is a long-standing concept in organisations and policing, there is currently no clear definition and operationalisation for the area of AI and internal security. AP4AI will move beyond existing approaches by establishing a coherent approach to achieving AI Accountability which is central to AI design, development and implementation. It will also be contextualised specifically for the complexities of operational practices in the internal security community.

AP4AI PRINCIPLES IN EXISTING DOCUMENTS

AP4AI puts forward 12 Accountability Principles that together define requirements for achieving AI Accountability in the internal security domain (see [section on AP4AI Accountability Principles](#) and [AP4AI Summary Report on Expert Consultations](#)).¹³⁴ The AP4AI Principles are based on existing work as reviewed above and refined through expert and citizen consultations (see [section on AP4AI approach](#)). The present section reflects on the grounding of the Principles in existing work to explain links, as well as highlight adjustments to fit them into the specific twin requirements of AI and its deployment in the internal security domain.

Legality

The need for application of laws and relevant regulations is widely accepted, in that any design, development and implementation of AI systems should be fully compliant with the relevant laws and regulations.¹³⁵ A recurring example to enshrine legality into AI systems is the concept of *privacy-by-design*. As emphasised by Cavoukian et al., “privacy-by-design and accountability go together in much the same way that innovation and productivity go together.”¹³⁶

Governing legal frameworks can define the objectives of governance and provide a legislative grounding for accountability practices,¹³⁷ with legislation being able to support determination who is legally responsible.¹³⁸ In some countries, there is a legal obligation for stakeholders to comply with the law for any AI system.¹³⁹ However, there is a dearth of uniform rules across different countries. An example

is the issue currently faced in the UK with respect to the lack of legal safeguards to follow when using AI systems, meaning that they are being developed to differing standards with differing levels of oversight and scrutiny.¹⁴⁰ The Law Society for England and Wales recommends that “the lawful basis of all AI systems in the criminal justice system must be clear and explicitly declared in advance.”¹⁴¹ However, with the lack of legal safeguards to follow, this may be difficult to achieve. It is crucial that GDPR and privacy laws are respected in AI, and assessments should be made in line with these through impact assessments.¹⁴² One way this could be achieved is through the use of Algorithmic Impact Assessments (AIAs), in line with the approach detailed in the EU’s proposed Artificial Intelligence Act.¹⁴³ The Ethics, Transparency and Accountability Frameworks for Automated Decision Making also states that approaches must be flexible to incorporate any changes to legislation or data. This is particularly prevalent in the field of AI, as legislation is constantly changing to keep up-to-date with technological innovation.¹⁴⁴ In the same regard, the Ada Lovelace Institute¹⁴⁵ states that legal frameworks are not enough on their own as their effectiveness depends on several factors, including political will and cultural norms. This may lead to hard legislation on AI actually impeding innovation.

Universality

Ensuring that all relevant aspects of the AI ecosystem and all actors involved in the deployment of AI within a specific context are covered within an approach has been lightly addressed in existing approaches. This principle was given significance in the recommendation to the Commission on Civil Law Rules on Robotics, whereby humanistic values that are prevalent to European societies should be reflected throughout AI design and development processes.¹⁴⁶ The ICO also supports this view, arguing that accountability is not restricted to the contents of the GDPR, but should be extended to all processing operations involving personal data.¹⁴⁷ This is also recognised with an ethical focus by the Alan Turing Institute,¹⁴⁸ which notes that ethical principles are anchored by a universal set of principles which focus on the “equal moral status” of humans in AI use. Although this is a specific area of AI without necessarily an operational context in mind, it can be more broadly applied to AI use in the internal security domain and highlights the importance of having a unified set of principles which are applicable to the entire AI system and associated actors.

In implementing Universality in AI, existing approaches have referenced the importance of oversight bodies in ensuring that all areas of the AI lifecycle are covered. For instance, the AI HLEG notes that regulation should use a harmonised approach that includes both implementation and enforcement mechanisms.¹⁴⁹ However, an accountability gap that needs to be addressed is that automated AI mechanisms are not themselves justifiable. Therefore, it must be ensured that individuals involved in AI use can be linked to decision making processes supported by AI systems.¹⁵⁰ The AI Now Institute explains this further, arguing that humans should be considered as part of AI decision-making processes, as they are responsible for classifying the input data, determining what goals the system should have, conduct system training and evaluations, and act upon the decisions and assessments made by AI systems.¹⁵¹

More specifically, scholars have discussed universality in relation to the policing context. For instance, Babuta and Oswald¹⁵² note that having universal evaluation standards is fundamental to ensuring the empirical validity and quality of AI systems used in the policing context. However, it is important that regulatory systems are not limited to only the police but include wider actors across the internal security domain. As emphasised by Zardiashvili et al.,¹⁵³ the use of AI is not limited to the police in isolation, but rather is used across the judicial chain including local governments and the judiciary for example. Thus, accountability processes should also consider external stakeholders which may have a role in the development or maintenance of AI systems with security applications.¹⁵⁴ AP4AI recognises this by implementing Universality as a multi-stakeholder approach which aims to ensure Accountability across all components and the complete lifecycle of the AI system.

Where existing approaches have referenced the notion of Universality to an extent, there is a lack of detailed discussion which covers specifically what Accountability should cover. For instance, existing approaches have not given explicit consideration to the processes beyond the criminal justice context, including design, development and supply which accountability should equally apply to. Furthermore, the approaches do not consider specifically which legal instruments for example should be covered under the principle. This is where AP4AI provides a significant contribution to ensuring that Accountability is achieved across the entirety of AI processes and mechanisms and provides examples of applicable laws which should be implemented.

Pluralism

Pluralism is described throughout the literature as a crucial aspect to achieving accountability. The inclusion of all relevant stakeholders ensures that no harm is omitted and allows for different perspectives to be assessed to ensure that there is no amplification of potential biases within the deployment of AI.¹⁵⁵ Pluralism enshrines the idea of multi-level participation which includes meeting the needs of affected communities with the aim to generate responses and increase trust within society if the oversight process prioritises public participation.¹⁵⁶ This will ensure that all perspectives are considered. As well as working to ensure accountability of AI practices, there is a need for more constructive and continuous multi-level collaborations to address ethical issues surrounding AI.¹⁵⁷ Pluralism thus explicitly emphasises the importance of avoiding homogeneity by only having regulators from the same background. Instead, regulators should include different actors such as civil society, public and private organisations, experts of fundamental rights, and should include those from under-represented and vulnerable communities.¹⁵⁸ The benefit of participants from different backgrounds is that they will inevitably provide a range of experiences and perspectives to ensure that biases are easier recognised and subsequently eliminated from any system results.

Transparency

Transparency is a principle that is documented across most approaches relating to AI. It is a common theme that holders of public office should act and take decisions in an open and transparent manner, and information should not be withheld from the public unless there are clear and lawful reasons for doing so.¹⁵⁹ While transparency is important in all AI systems, the nature of the work within the internal security domain and the potential high-risk outcomes from AI systems make it even more important that systems and related processes and decisions can be viewed. More specifically, Transparency can help to resolve the questions of responsibility and liability within expert debate.¹⁶⁰ Therefore, Transparency is vital in ensuring trust and determining who or what is accountable for potential problems with AI systems.¹⁶¹

As found across the literature which explores public perceptions of general AI and its use, experts have concluded that public understandings are “broad” but not “deep.”¹⁶² This is emphasised by the Centre for Democracy and Technology, which notes that “cultural perceptions of automated decision-making technology are out of step with the technical reality.”¹⁶³ This is perhaps due to the rise in systems which are increasingly autonomous and “invisible”, which becomes difficult for the public to scrutinise.¹⁶⁴ This is reflective of the Information Society being redefined as a “black box society”, whereby algorithms are difficult to read and lack legibility.¹⁶⁵ The principle of Transparency is therefore critical in closing the knowledge gap within the public and ensuring that the AI actions being undertaken within the internal security domain are given trust and confidence.

In order to mitigate against these concerns, Bristows notes that the AI industry “should be accountable and responsible to the public.”¹⁶⁶ Specifically as AI use can have direct implications on communities, it is important that they are aware of the decision-making processes and mechanisms as well as the impact this can have.¹⁶⁷ Therefore, in the context of the internal security domain, ensuring transparency is a key pillar to achieving accountability and building public trust in AI use in the security domain. As emphasised in the UK Digital Policing Strategy 2020-2030, “appropriate and transparent consideration of ethics in pursuing these priorities is critical to maintaining the integrity of our policing service and the trust of the public.”¹⁶⁸ Although this has an ethical focus, it outlines the importance of Transparency within the context of modern policing.

It is essential that in order to build trust, increase transparency and minimise the risk of bias or error, AI systems are developed in a manner which allows humans to understand their actions.¹⁶⁹ There are however arguments that as AI tools become more sophisticated, they also pose real threats to the transparency and democratic accountability of practitioners in the area of internal security.¹⁷⁰ Some particularly complex methods of big data analysis, for instance, can make it difficult for organisations to be transparent about the processing of personal data.¹⁷¹ These types of ‘black box’ systems rely on the feeding of information through many different ‘layers’ of processing in order to come to an answer or decision, making the potential for full transparency increasingly difficult.¹⁷² Therefore, frameworks must include Transparency mechanisms which ensure that it is understood how AI can infringe upon fundamental rights, for example.¹⁷³

However, there is a caveat to Transparency which must be considered in the context of the internal security domain. A particular issue faced by internal security practitioners is the potential requirement for some sensitive policing information to remain hidden. In such cases it is sometimes undesirable to achieve full transparency.¹⁷⁴ Providing full transparency is not always productive, and in the case of security-related uses can even be dangerous, as it allows bad actors the potential to exploit or circumvent the AI systems being deployed.¹⁷⁵ Interestingly, however, studies have shown that the public recognises this and argues that police should not always be transparent. For example, the Britainthinks report¹⁷⁶ explored public perceptions of AI use from a policing use case about neighbourhood policing. The participants argued that in this scenario making the decision-making processes publicly available could enable potential criminals to take advantage of the information. In a specific case, a participant emphasised that there must be mechanisms in place to determine whether the information was deemed as safe for public view.¹⁷⁷ Similar results are found in the AP4AI citizen consultation with a specific focus on AI Accountability (see [section on Citizen consultation](#)). This aspects highlighted in the AP4AI Framework, which recognises that Transparency must be achieved in a ‘timely, meaningful and appropriate way’, that considers the operational context which internal security practitioners face in ensuring accountability.

Independence

The principle of Independence in the AP4AI framework refers to the status of competent, independent authorities performing oversight functions in respect of achieving accountability. While Independence was not widely cited within the literature as a specific principle, a number of authors referred to the need for independent oversight bodies. It has been stated that independent oversight bodies should ensure accountability through the monitoring of actions of those designing and implementing AI systems and reviewing related processes.¹⁷⁸ Oversight bodies may also be used as a platform where expertise in the area is consolidated, allowing for the effective application of accountability mechanisms.¹⁷⁹

Within the literature, there is reference to the need for legal issues to be considered by oversight bodies, with some stating that AI specialists are needed to create a legal framework to ensure that experts within these independent oversight bodies are monitoring the legal and ethical implementation of AI.¹⁸⁰ Further, it was suggested that the oversight of AI could come in the form of either legislation or advisory capacities.¹⁸¹ While the literature does mention the need for oversight to be conducted independently, there is no specific mention to who should be conducting such a function. The AP4AI definition draws upon the need for the oversight body to be independent in every way, ensuring there is no conflict of interest in any sense. This will support credible reviews of the functioning of the AI systems, its process and impacts.

Commitment to robust evidence

The AP4AI Principle of Commitment to Robust Evidence demonstrates and facilitates accountability by requiring detailed, accurate and up-to-date record keeping in respect of all aspects of AI use. This principle is included in the AP4AI Framework, as the production of empirical evidence is key to accountability, ensuring that a system's performance can be adequately assessed.¹⁸² The literature on this topic demonstrates how the principle of Commitment to Robust Evidence links heavily to other principles included within the AP4AI Framework. For instance, if an AI system is not explainable, this can cause difficulty in interpreting the results and outputs,¹⁸³ hindering the possibility of obtaining robust evidence.

The appointment of oversight bodies, as detailed in the principle of Independence, will have a significant effect on the quality of evidence produced from the AI systems, as these bodies would have the ability to establish monitoring systems to evaluate and identify issues in the AI performance and effects before this is presented as evidence.¹⁸⁴ These bodies should also set oversight mechanisms which allow for the systems to be scrutinised and ensure any potential risks are mitigated.¹⁸⁵

Although this principle is not heavily referenced within the literature, it is a crucial step to ensuring the quality of evidence presented from and about an AI system upholds the standards of prosecution evidence in terms of integrity, credibility and continuity. In this regard, a documentation and evaluation process, as part of this principle, are key to good performance, robustness, security and safety of the AI systems.¹⁸⁶

Enforceability and redress

Both Enforceability and Redress are extensively covered in the literature as essential components to ensuring accountability is upheld when deploying AI systems. A key element in existing discussions is the need for effective governance mechanisms to oversee that AI systems used by internal security practitioners are necessary, proportionate and lawful, and that AI systems are designed in ways that help to mitigate any risks identified.¹⁸⁷ It is recommended that given the speed of development and implementation of AI, a regulatory assurance body should be considered, who can identify gaps in the regulatory landscape and provide advice to individual regulators and government on the issues associated with AI.¹⁸⁸ The AP4AI definition of Enforceability and Redress states that mechanisms should be established that facilitate independent and effective oversight in respect of the use of AI in the internal security community. This is reiterated in the ethics framework of Australia, in which the Law Council recommends that an ethical and regulatory framework be implemented formally so as to provide for enforceability.¹⁸⁹

The principle of Redress and the need to provide effective remedy to those who have been wronged by AI systems and their outputs is something that is agreed upon as a necessity for any real accountability. To remain accountable for their decisions, public bodies need to enable people to challenge decisions and to seek redress using procedures that are independent and transparent.¹⁹⁰ In the context of the EU, it has been argued that EU Member States need procedures in place

to ensure that those who might be negatively impacted by AI systems have an effective and accessible remedy against those responsible,¹⁹¹ as public bodies and those carrying out public functions have to act in accordance with public and administrative law principles and must act lawfully, rationally, proportionately and fairly.¹⁹² An effective means of ensuring redress as a feature of accountability within AI systems is to incorporate a redress-by-design mechanism.¹⁹³ This will provide those developing and implementing AI systems under the AP4AI Framework, with a means of ensuring that those affected by AI decisions are adequately protected from the outset.

Compellability

The principle of Compellability is not expressly described in the literature as a standalone concept or principle in its own right, but often ties into discussions surrounding oversight bodies. The AP4AI Framework includes Compellability in its own right as a means of giving oversight bodies the power to compel those organisations 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. This was deemed necessary to ensure that provisions can be put in place to compel organisations dealing with AI to provide the necessary information to allow meaningful AI Accountability.

Explainability

The necessity to have explainable AI is featured across the relevant literature. It has also been identified as part of other principles such as Transparency. However, AP4AI argues the importance of Explainability as a standalone principle. Explainability does have links to the principle of Transparency, as it provides an example that, while the default position should be full transparency, appropriate alternatives that achieve the same aim may be implemented, in cases where legal or sector-specific constraints apply or in relation to the use of Blackbox AI tools, which are inherently opaque.

To make AI accountable, decision makers must be able to justify the outputs of systems.¹⁹⁴ The concept of Explainability therefore obligates AI systems and the organisations deploying them to supply evidence, support or reasoning for each output, which should be tailored to the understanding of different stakeholders at five levels: user benefit, societal acceptance, regulatory and compliance, system development and owner benefit.¹⁹⁵ If an AI system is explainable, this will lead to an increase in trust amongst users and the public¹⁹⁶ and provide support should a dispute arise.¹⁹⁷ Most of the literature surrounding explainable AI is in line with the AP4AI Principle, ensuring that actors at every level can understand the AI system itself, its potential effects, and subsequent resulting effects. Also mentioned within the literature is the importance of explanations being delivered through mechanisms that are accessible and understandable,¹⁹⁸ which is integrated as a core feature of the AP4AI Explainability principle.

In discussion of Explainability, existing approaches have often referred to potential issues which can arise with the use of AI. A particular concern raised are potential biases in AI systems, which have been widely documented.^{199,200} Recognising such biases within complex algorithms is not straightforward and will require experts to monitor and assess the decision-making processes.²⁰¹ Therefore, where possible and in order to reduce the potential bias in the system and its results, it is argued that AI should be designed and developed in an understandable manner.²⁰²

In a similar vein to Transparency, experts have also noted the issues with private organisations being too open in their explanations which can provide details of their systems to competitors.²⁰³ This can be contextualised to the internal security domain, whereby providing explanations must also be sensitive to compromising policing methods and their effectiveness. This raises the question of “how much of a system/algorithm can be explained to users and stakeholders?”²⁰⁴ As a solution, the Explainability principle should have a clear focus on both the process and results.²⁰⁵ This will ensure that stakeholders are provided with a general explanation of the decision-making process and what factors were included in the decision.

In developing Explainability, relevant stakeholders will be able to understand how algorithms reached a certain result, which will provide a space for checks to be conducted into potential areas of concern which could not be challenged without explainable AI providing evidence of how the system came to its output.²⁰⁶ For example, existing approaches have outlined that Explainability can be achieved through technical and regulatory audits. Specifically, technical audits can be beneficial in testing the inputs and outputs of AI systems to examine if there are signs of racial bias, for instance.²⁰⁷ The use of auditing is also supported by Raji et al., who argue that this can be fundamental to closing the accountability gap in developing and deploying AI systems.²⁰⁸ Therefore, ensuring Explainability of the output is crucial in avoiding errors and increasing trust in the system.²⁰⁹ However, scholars have argued that existing approaches have failed to account for what should happen after an organisation has been made to explain their AI systems, mechanisms or processes; in other words, “what kinds of accounts should we accept as valid?”²¹⁰ AP4AI aims to address this by ensuring that practitioners consider how effective Explainability is determined and whether there are review mechanisms in place to scrutinise an explanation given.

While some aspects of the literature place an emphasis on the need for Explainable AI in order to build trust in the system, the AP4AI Principle does not explicitly include this wording. It is also stated that the requirement of consistent Explainability may be too heavy a burden for organisations operating AI systems, particularly when considering the use of Blackbox AI.²¹¹ This is supported by the *Partnership on AI*, who argue that Explainable AI is often regarded as the solution to understanding the Blackbox and how predictions are made. Despite this, they stress that existing approaches do not adequately enable practitioners to make effective and meaningful explanations.²¹² Therefore, the Explainability principle under the AP4AI Framework aims to address this and ensure that Accountability is enhanced through effective Explainability techniques and best practices.

Constructiveness

The principle of Constructiveness is widely referenced in the literature as an important means for ensuring AI is developed and used in a way which allows for engagement at every step. The AP4AI definition of Constructiveness encompasses 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 includes that individuals who are subject to AI have the ability to complain and receive remedy in case of effect (i.e., linking Constructiveness to Enforceability and Redress), with the avenues for appeal being openly presented to them.²¹³

Along with the inclusion of Constructiveness in some AI frameworks, major companies who use AI have set out how they will ensure Constructiveness in its design and implementation. Google, for instance, stated that their AI will be subject to human direction and control, providing those affected with opportunities for feedback, explanations, and appeal.²¹⁴

As stated in the AP4AI definition, it is important that effective constructiveness comes from multi-level participation with different actors in the assessment of AI, and that the general understanding of AI usage by security practitioners is openly accessible, ensuring that individuals who are not technically minded are also able to interpret the findings.²¹⁵ This will allow for the constructive dialogue to be developed by a democratic and multi-participatory body.

The AP4AI Principle of Constructiveness is very similar to the current definitions provided in the literature. However, the AP4AI Framework aims to build upon the present definitions by implementing mechanisms within the principle, which provide users with opportunities for feedback, explanation and appeal. This goes beyond the state of the art, ensuring that practical steps are included to achieve true constructiveness in the design, implementation and review of AI for security and policing and the AI Accountability process more generally.

Conduct

The principle of Conduct, as put forth in the AP4AI Framework, is not discussed in the literature as specifically relating to conduct when using AI systems. Conduct aims to sit alongside other principles within the Framework to ensure that not only is there overall accountability for organisations using or developing AI, but that individual conduct is also held to account. As mentioned in the meaning of the principle, the European Code of Police Ethics states, “the condition of a democracy can often be determined just by examining the conduct of its police.”²¹⁶ Conduct features heavily across national policies as a requirement for LEAs when conducting any investigation, and its inclusion in the AP4AI Framework aims to ensure that this is the case when any investigation or other activity includes the use of an AI system in the broader internal security domain.

Learning organisation

It is crucial with every new technology that organisations are willing to take on a role of continuous learning to ensure the application of new knowledge and insights, as detailed in the AP4AI definition. This sentiment is echoed across the literature, with the Council of Europe recommending the promotion of AI literacy by the government, oversight bodies, human rights structures and the judiciary, to facilitate the advancement of knowledge and understanding of AI.²¹⁷

It is moreover widely recognised that due to the substantial skills shortage in AI at present, more training is required.²¹⁸ This should be continuous efforts, as without the correct knowledge, the potential of biases is amplified.²¹⁹ Along with continuous training, the knowledge of the AI systems within the internal security ecosystem should also be subject to continuous improvement through education and information exchange.²²⁰

In order to deliver effective training within AI, an interdisciplinary approach is required.²²¹ Governments and oversight bodies will benefit from collaboration and the ability to share best practices across stakeholders that arise when implementing accountability mechanisms.²²² The AP4AI principle of Learning Organisation is pivotal to the whole of the framework, as the use of AI is a continuous learning stream, and without proper training and investment in staff, it is difficult to effectively implement and assess AI Accountability.²²³ It is recognised that training must be designed in a way that serves different actors²²⁴ based on their roles and knowledge, and that the mechanisms for training are successfully able to build awareness, inform, prepare and upskill the stakeholders.²²⁵ While the literature provides a number of similarities to the AP4AI definition, there is no explicit reference to the need for improvement and modification of systems within the continuous learning of an organisation. This is a key part of the AP4AI Framework to ensure that learning is facilitated across both the people and systems of the entire internal security ecosystem.

LEGAL FRAMEWORKS WITH RELEVANCE FOR AI AND APPLICATION IN AP4AI

In addition to the considerations and frameworks discussed above, the AI domain also engenders intense ongoing debates and efforts with respect to legislation. This section details core legal frameworks relevant to AI, which underpin and inform efforts in AP4AI.

A starting point to consider are the Digital Service Acts (DSA) 2000 and 2020 created by the European Union to harmonise Fundamental Rights and establish a “level playing field” in businesses.²²⁶ The DSA 2000 was first introduced on the 17th of July 2000 and designed to meet the request of the European Union to “create a safer digital space.” Since its first publication, several amendments have been made which have brought the act up to date. The importance of this act for the current context is its reference to ‘responsibility and accountability’.²²⁷ The separation between what can be classified as ‘accountable’ and what is classified as ‘responsible’ shows that there is a call for a systematic understanding of how accountability is defined. Given the current lack of a clear accountability definition for AI, as well as for AI deployment by the internal security domain, it also highlights the need for a specific framework that details how accountability can be defined, operationalised and achieved in this area.

More recently, the Digital Market Act created in 2020 was designed to harmonise the large-scale digital services that are spread across the European Union Digital Market. The Digital Market Act does not refer to any accountability structures or requirements as requested by the DSA. Rather the DMA is designed to protect businesses to ensure that large providers of core platform services do not gain significant access to large amounts of data. This act is relevant to AP4AI as it creates the environment upon which LEAs and other internal security actors may need to deploy AI solutions for protection, identification of threats, safeguarding or the prevention of illegal activities.

In April 2021 a significant change in the way AI was perceived by the Member States of the European Union was achieved with the proposed Harmonisation of Rules on AI (*Artificial Intelligence Act 2021*).²²⁸ This bill is designed to introduce a “wide array of economic and societal benefits across the entire spectrum of

industries and social activities.”²²⁹ The act also highlights new areas of inquiry, in that “AI systems will be required to be in line to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules.”²³⁰ The *Artificial Intelligence Act* is set to affect how AI stakeholders respond to new state-of-art technologies within the digital industry and will thus also need to inform the final AP4AI Framework and its application.

Another relevant document is the General Data Protection Regulation (GDPR) or Regulation 2016/679.

GDPR is paramount in referring to the protection of data which is a fundamental element in the practical application of AP4AI for concrete AI deployments to protect the rights of individuals by ensuring that the highest quality of AI-based technology is implemented by actors in the internal security domain. This refers specifically to the handling of data through AI using automated decision-making as discussed under the GDPR Article 22 of Regulation 2016/679 (referring to profiling and the legal implications of automated systems which are concerned with ethical matters of protecting an individual’s rights).

Article 22 ensures that data controllers are aware of human involvement when referring to AI. This refers to the processing of a data subjects’ information which would create a significant impact upon them. The data controller must consider legitimate interests and safeguard their rights and freedoms which should be paramount to the design and deployment of AI. The data controller must also understand the concept of ‘consent’ when referring to decision-making by AI. In most cases, consent will be required by the user unless a case can be provided to show that there is a strong significant interest, and the rights of an individual are protected. This article also refers to the authorisation within law. The latter gives a legal basis to the AP4AI Framework to focus on the legalities of AI during the deployment of any AI capability within society by internal security practitioners (see [section on AP4AI Framework Blueprint](#)). These factors are fundamental to the AP4AI Principles (specifically Legality), as well as its application, as it provides important framing towards positive data handling mechanisms (also see Directive (EU) 2016/680).²³¹

Relevant to AP4AI is also the concept of a ‘Right to Explanation’, which is implied in the Regulation under Article 14(2)(g) which states that:

“In addition to the information referred to in paragraph 1, the controller shall provide the data subject with the following information necessary to ensure fair and transparent processing in respect of the data subject: (g) The existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.”²³²

While not directly mentioned, Regulation 2016/679 thus requires elements of explainability and assessment to certify that good practices have been complied with. AP4AI explicitly integrates Explainability, together with Transparency, as part of its principles, in recognition of the importance that access to appropriate information plays for accountability.

While the above discussion focused on European efforts, AI is of course also a substantive area for regulation on a global scale. Below we detail a select number of countries, focusing on the United States, Canada, Australia and the UK, as far as they relate to accountability and interact with AI regulations, guidelines or frameworks.

The United States has a number of regulations that may help inform a universal Accountability Framework. In California, the *Automated Decision Systems Accountability Act* refers directly to methods of accountability in AI and Machine Learning (ML).²³³ A similar approach referring to accountability exists in New York, which passed the *Algorithmic Accountability Act of 2019*.²³⁴ These acts incorporate a similar approach to the European Union by utilising Impact Assessments.

In Canada, areas of law which relate to accountability have been equally at the forefront of discussion. As Thomassen states, “negligence and strict liability claims are likely to be more common legal mechanisms in the AI context.”²³⁵ This perspective highlights the understanding that liability and legal frameworks about who can be held accountable stem from areas of tort law under negligence. The Ethics Framework proposed by the Law Council in Australia, for instance, comprises principles to be followed when introducing AI in an operational manner, namely: (a) generation of net benefits, (b) doing no harm, (c) regulatory and legal compliance, (d) privacy protection, (e) fairness, (f) transparency and Explainability, (g) contestability and (h) accountability.²³⁶

These perspectives benefit the implementation of AP4AI, as methods within the principles can mitigate against creating loss or harm to the parties involved when internal security practitioners deploy AI in their respective countries. The overlap in principles across national discussions and AP4AI suggest the existence of a robust set of factors that can result in a strong and future orientated framework.

In the United Kingdom, an extensive *National AI Strategy*²³⁷ was created by the Central Digital and Data Office designed to ensure that AI is being appropriately utilised, and the risks are dealt with accordingly. The *National AI Strategy* was produced in reaction to a call by the UK government designed to build “on the UK’s strengths but also represents the start of a step-change for AI in the UK, recognising the power of AI to increase resilience, productivity, growth and innovation across the private and public sectors.”²³⁸ The *National AI Strategy* sets out three pillars: (a) long-term needs of the AI Ecosystem, (b) ensuring that AI benefits all sectors and regions and (c) governing AI effectively. The *National AI Strategy* supports research innovations in the UK for AI usage and for finding new legislation to determine the rules on AI. Related to the *National AI Strategy* is the *Algorithm Transparency Standard (ATS)*.²³⁹ The ATS has been designed to enforce “research that will help develop a cross-government standard for algorithmic transparency.”²⁴⁰ The ATS addresses areas with relevance for the application of AI such as technical specifications, potential public effects (e.g., has a potential legal, economic, or similar impact on individuals or populations, affects procedural or substantive rights, affects eligibility, receipt or denial of a programme) and impact on decisions (e.g., replaces human decision-making, assists or adds to human decision-making).

The above is only a limited extract of existing legal discussions and regulations in the context of AI. The short overview highlights, however, the relevance of accountability also in the legal domain, and in the same regard the specificity of the legal bases in case of specific AI deployments. Moreover, the ability to apply existing legal frameworks, ethics and human rights concerns creates a significant challenge. This means that, overall, Legality is a universal concern, but that relevant laws have to be considered per context.

LIABILITY IN THE CONTEXT OF ARTIFICIAL INTELLIGENCE

The AP4AI Framework ensures liabilities and accountability are in the focus of AI deployments. Therefore, an important strain of the law AP4AI may also turn to is the legal concept of Tort Law and Contract Law. The core concept of these two areas is the element of liability to others. In Tort Law a breach of a duty owed to an individual – direct vicarious – can lead to legal pressure. This applies similarly to contract law, when referring to a breach of contract, for example, if the purported AI was poorly designed or failed to mitigate data protection losses or if it otherwise failed to meet the terms as agreed within any procurement or other enforceable agreement. This argument has been supported within the United Kingdom where the UK Government stated that “an individual claimant could seek to obtain a remedy from the UK courts in relation to a certain biased or heavily skewed outcome of an algorithmically-based decision-making process.”²⁴¹ Hacker and colleagues in their paper ‘Explainable AI Under Contract and Tort Law’ further argue, that Explainability “crucially influences questions of contractual and tortious liability for the use of ML [machine learning] models.”²⁴² The importance of these two linked areas – Explainability, as well as the option of liability – is reflected in the AP4AI Framework under the Principles Legality, Explainability and Enforceability and Redress (see [section on AP4AI Accountability Principles](#)).

AI AND FUNDAMENTAL RIGHTS

“Virtually all human rights can be affected by the use of AI systems. Various actions are therefore needed, amongst which: thorough assessments of the effect of AI systems; independent and expert scrutiny; transparency on the use of AI; ensuring the availability of remedies; new legal frameworks to codify the principles and requirements governing the use of AI, in conjunction with voluntary ethics codes committing AI developers to act responsibly.”²⁴³

This statement made by the CCBE highlights the call for a systematic consideration of Human Rights in the context of AI.

Fundamental Rights and citizens’ access to these rights are crucial in the application and assessment of AI, and especially in achieving the vital balance of “protecting basic human rights while fostering innovation.”²⁴⁴ In achieving this balance within the internal security domain, it is important to identify any positive obligations on the State to protect citizens²⁴⁵ and the extent to which their ability to meet those obligations would be enhanced by AI. The European Union Agency for Fundamental Rights (FRA) has provided extensive insights and recommendations

surrounding AI and the rights of citizens, also with explicit reference to the law enforcement domain. For instance, while focusing on facial recognition, FRA highlights that “full compliance with fundamental rights is a prerequisite for any law enforcement activity, irrespective of the technologies used.”²⁴⁶ Specifically, FRA points to the importance of avoiding poor data quality,²⁴⁷ discrimination and biases,²⁴⁸ unlawful profiling²⁴⁹ and access to remedy and complaint mechanisms.²⁵⁰

The FRA further details that when introducing new legislation relating to AI or any form of new policy, “relevant safeguards need to be provided for by law to effectively protect against arbitrary interference with fundamental rights and to give legal certainty to both AI developers and users.”²⁵¹ This is reflected in the new *Proposal for a 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*, which requires the presence of a new instance to assist the act in providing Human Rights Due Diligence. Generally, FRA calls for human rights to be an integral part of any AI design and deployment and regular assessments of its impacts.²⁵² AP4AI fully endorses this perspective and is committed to ensuring a solid grounding in Fundamental Rights as enshrined by relevant frameworks.

Relevant passages of the EU Charter of Fundamental Rights highlight implications surrounding dignity, freedoms, equality, solidarity, citizen rights and justice. The Charter details the specific rights that link to AI, understanding of which creates important foundations also for AI applications in the internal security domain. Generally, AI deployments should respect all titles found within the Charter; however, specific areas have particular bearing for AI in the security domain.

Ensuring that Human Dignity is respected while developing and deploying AI requires concerted scrutiny to understand how to avoid infringing upon the basic characteristics of an individual. *Title I* is paramount as it ensures that individuals are not subjected to harm because of technology. Respect for Private and Family Life and the Protection of Personal Data within *Title II* are equally crucial in an AI context as core elements to safeguard against infringing on individuals’ rights, either by intent or negligence. More specifically, the Respect to Private and Family Life refers to privacy in online and offline realms, and AI should adhere to this principle by design.

The Fundamental Rights Charter also highlights key rights of citizens in the area of Equality. A significant issue related to this right is the potential of bias and discrimination in AI, which has been at the forefront of academic and legal discussions. Issues such as algorithmic bias, defined as a “systematic and repeatable error in computational systems, that is responsible for unfair, wrongful results of data processing,”²⁵³ requires careful identification and assessment, which AP4AI aims to support throughout the AI lifecycle. Titles relating to Fundamental Rights is *Title V* on Citizen Rights. *Title V* contains the power of citizens for free movement to the right to hold votes and petition. In all these areas, AI could affect the standing of Human Rights.

Title VI Justice is especially significant in the context of AP4AI and AI Accountability. *Title VI* calls for the right to fair trials, principles of legality and the rights to criminal proceedings. This principle is enshrined in AP4AI through Legality, ensuring the

compatibility of the AP4AI Framework. Finally, *Title VII* as the *General Provisions* of the EU Charter of Fundamental Rights are to be followed and complied with to ensure that the desired scope is appropriately satisfied.

The current discourse on Fundamental Rights and AI contains a mixture of opinions and recommendations. The newly proposed *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act 2021)*²⁵⁴ has resulted in several discussions regarding the position of Fundamental Rights within the new act. The *Artificial Intelligence Act (AIA)* does not currently include a clear Fundamental Rights position. Therefore, NGOs such as Access Now have set out their goals of what they propose for a new AIA. The suggestions have been organised into a Civil Society Statement, which requests the realisation of nine principles to ensure that Fundamental Rights underpin the implementation of the legislation into society.²⁵⁵ A number are directly relevant for AP4AI, as they offer concrete pointers for applying Human Rights in an AI context and are cited below for reference.

- **A cohesive, flexible, and future-proof approach to ‘risk’ of AI systems:** The concerns produced in this section refer to the wording found in Proposal 2021/0106 which does not contain specifics regarding the ‘risk’ of AI deployments. It has been found that this classification “does not consider that the level of risk also depends on the context in which a system is deployed and cannot be fully determined in advance.”
- **Prohibitions on all AI systems posing an unacceptable risk to fundamental rights:** This aspect explicitly proposes that “some AI practices are incompatible with EU rights, freedoms and values, and should therefore be prohibited.” This proposition relates also to future AI capabilities and how citizens may interact with new technologies. The report suggests that a “robust and consistent update mechanisms for unacceptable and limited AI systems” should be introduced. There is also an overview of vulnerabilities that could threaten Fundamental Rights with respect to specific AI applications that are of direct relevance for the internal security domain: (a) “the use of AI systems by law enforcement and criminal justice authorities to make predictions, profiles or risk assessments for the purpose of predicting crimes; (b) The use of biometric categorisation systems to track, categorise and / or judge people in publicly accessible spaces; or to categorise people on the basis of special categories of personal data, protected characteristics, or gender identity.”
- **Obligations on users of high-risk AI systems to facilitate accountability to those impacted by AI systems:** The AIA aims to improve risk-mitigating efforts for high-risk AI and the enforceability of the AIA. Hence, there should be an “obligation on users of high-risk AI systems to conduct a fundamental rights impact assessment (FRIA).”²⁵⁶
- **Improved and future-proof standards for AI systems:** AI as a fast-developing area needs to ensure that “harmonisation under the AIA is without prejudice to existing or future national laws relating to transparency, access to information, non-discrimination or other rights, in order to ensure that harmonisation is not misused or extended beyond the specific scope of the AIA.”
- **Truly comprehensive AIA that works for everyone:** The main requirements for an act that satisfies Fundamental Rights are formulated as: (a) ensure data protection and privacy for persons with disabilities, (b)

ensure that privacy and data protection of all persons, including those under substituted decision-making regimes such as guardianships, are protected when their data are processed by AI systems and (c) financial implications of the AIA must be reassessed and planned so as to ensure that enforcement bodies and other relevant bodies have the resources to meaningfully fulfil their tasks under the AIA. The latter is of particular interest to AP4AI as practice-oriented Framework, as an Accountability mechanism is only feasible given sufficient resourcing and capabilities.

The above observations have given rise to discussions on '*How to Fix the EUs Artificial Intelligence Act*'.²⁵⁷ An instance is the call for "regulatory limits" on the use, as "without appropriate limitations on the use of AI-based technologies, we face the risk of violations of our rights and freedoms by governments and companies alike."²⁵⁸ Amongst the greatest concerns for AI deployments is the lack of transparency, meaning that society has an interest in the Explainability of AI practices. These discussions also reference the issue of how to install an "AI ecosystem of trust and excellence", generally proposing meaningful transparency and accountability for how AI is developed, marketed and deployed.²⁵⁹ These elements also emerged prominently in the AP4AI expert consultations²⁶⁰ and were integrated into its Accountability Principles.

The Council of Europe argues that laws and regulations can be deemed to be limited in their "applications of AI that are incompatible with fundamental rights."²⁶¹ It also highlights specific issues in this regard:

1. Prohibited practices are too vague
2. Many practices currently labelled "high risk" need to be prohibited
3. Lack of criteria for prohibited practices

In reaction, it proposes concrete recommendations for Prohibited Artificial Intelligence Practices:

1. Uses of AI to categorise people on the basis of physiological, behavioural, or biometric data, where such categories are not fully determined by that data
2. Uses of AI for emotion recognition
3. Dangerous uses of AI in the context of policing, migration, asylum, and border management

A commonly proposed method to this end is that "all users of high-risk AI systems should be obliged to perform either a data protection impact assessment (DPIA), or, where a DPIA is not applicable, they should be required to carry out a human rights impact assessment (HRIA)."²⁶² This assessment of Human Rights is considered as a "necessary requirement in the development and deployment of AI solutions to prevent any prejudice to human rights and fundamental freedoms, as well as to promote a human rights-oriented AI"²⁶³ (cp. [section on AP4AI Framework Blueprint](#)).

Human Rights Impact Assessments

The concept of Human Rights Impact Assessment (HRIA) has been created by the United Nations, designed to ensure that Human Rights due diligence is accomplished. The *Guiding Principles on Business and Human Rights*²⁶⁴ detail how businesses and infrastructures should integrate Human Rights into their structures. This document defines HRIA as a “process to identify, prevent, mitigate, and account for how they address their impacts on human rights [... and] to enable the remediation of any adverse human rights impacts they cause or to which they contribute.” Additional guidance is provided by the Danish Institute for Human Rights,²⁶⁵ While Data & Society offer a targeted approach for an AI-based implementation of Human Rights Impact Assessment.²⁶⁶ These approaches ensure that human rights are developing into the “the core of future AI regulation.”^{267,268} and will be consulted intensely in the further development of the AP4AI Framework.

CITIZEN CONSULTATION

Contextualisation is vital to give insights into the cultural, social and political values that determine which principles and implementation processes are meaningful for AI Accountability in the internal security domain across operational and national contexts.^{269,270, 271} Contextualisation cannot only help identify which key values and priorities are shared across contexts but also identify variations and specifics. This step is therefore core to achieving AP4AI's ambition of creating practical mechanisms and tools that directly and meaningfully support AI Accountability.

The best way to understand contextual differences is broad stakeholder engagement across disciplines and national contexts. AP4AI does so in an integral way throughout all its activities using an expert-driven approach (see [section on AP4AI approach](#)). Expert insights are crucial for the development of any framework impacting the use of AI by security and policing agencies to guarantee it incorporates the opinions and knowledge of experts within the relevant fields. Consultations with experts and stakeholders are a key driver for the development of AI strategies²⁷², and subsequently any relevant frameworks.^{273,274}

AP4AI has conducted a broad consultation with subject matter experts in Cycle 1, results of which are reported in the [AP4AI Summary Report on Expert Consultations](#).²⁷⁵ Yet, for AP4AI core expertise also lays with members of the public that are directly affected by AI deployments by security practitioners. Further, citizens are a core stakeholder to accountability in the security domain. AP4AI is therefore conducting a second consultation with citizens across 30 countries (see details below).

The triangulation and integration of multi-sectoral and citizen perspectives ensures that the AP4AI Framework comprehensively combines LEA, legal, human rights, ethical, technical and citizen perspectives, and is cognisant of contextual variations. Engagement with citizens reacts to requests that “policies should prioritise public participation as a core policy goal.”²⁷⁶ A second benefit of consultation and engagement with citizens is that it supports the creation of trust and legitimacy within communities by giving them a voice in the creation of AI Accountability mechanisms. Incorporating citizen perspectives can validate and solidify an approach that reinforces public confidence in the authority's adherence to organisational accountability as the precondition for AI deployments. As emphasised by Haataja et al. in outlining the AI register, civil participation can be invaluable in ensuring the public has an input on the impacts of AI within their communities.²⁷⁷

APPROACH

The citizen consultation is the second consultation cycle within AP4AI after the expert consultation (see [section on AP4AI approach](#)). The purpose of the citizen consultation is to validate and refine the set of 12 AP4AI Principles developed in Cycle 1²⁷⁸ from a citizen perspective and obtain insights into expected and accepted AI Accountability mechanisms.

As the principal group in any democratic policing and justice model, the consultation and engagement with citizens is a core milestone within AP4AI and a vital element of the project's expert-driven approach. If citizens, in whose name security measures are conducted, are not involved in a meaningful way, any framework claiming to enhance democratic accountability lacks structural credibility.

The consultation targeted adult participants (18+) from the general population in 30 countries (27 EU Member States, Australia, UK and USA) to obtain a broad and varied sample within and across countries (i.e., no demographic group, profession, etc. or were specifically targeted or excluded).

Due to the scale of the engagement, the citizen consultation was conducted as an online survey.²⁷⁹

The surveys were presented in the respective country language(s) to ensure that participants could answer questions without language barriers.

All participants were asked to give their informed consent before answering the survey.²⁸⁰ The study received ethics approval by Sheffield Hallam University, as home institution of CENTRIC, which leads the empirical activities in AP4AI.

Aspects addressed in the consultation

Understanding citizens' perspective on AI Accountability was the core purpose of the citizen consultation. To achieve this aim, the consultation captured citizen perspectives around the following four themes:

1. General attitudes towards AI use by police
2. Relevance of AI Accountability and accountability mechanisms
3. Reactions to the initial set of AP4AI Principles
4. Responsible actors for AI Accountability

The consultation further allowed participants to provide recommendations for additional accountability aspects and general comments or suggestions. The questions themselves are presented in the finding section.

The survey focused on 'police' instead of 'internal security domain' in all explanations and questions. This focus was chosen explicitly, as the term 'internal security domain' is highly abstract and likely to confuse citizens. In contrast, citizens are likely to have concrete views about police as most visible representative of the security practitioners.

Sample description

The citizen consultation will collect 6,400 answers across the 30 countries. At the time of this report, 5,239 answers were received, which form the basis of this report. The consultation continues until the full set of participants is collected, and findings from the complete sample will be described in a subsequent report.

Table 1 presents characteristics for the full sample. The gender and age distributions are balanced across categories (based on pre-determined quotas to reflect population characteristics), while the participant pool includes a good spread across educational levels. The sample further includes about 10% of participants, who describe themselves as member of an ethnic minority in their country, while about 33% have past experience with crime. The sample also contains participants with a security-related profession (9.6%). The self-ascribed knowledge about AI indicates moderate expertise ($m=3.26$, $sd=.89$), while on average participants rated their expertise about AI use by police as moderate to good ($m=3.80$, $sd=.90$). Overall, the sample characteristics demonstrate that the citizen consultation as intended managed to engage a broad and diverse set of participants.

Table 1: Characteristics for the full sample

Country	Sample Size	Gender Distribution*	Age Distribution*	Highest Education	Security-related Work	Ethnic Minority	Crime Victim
All Countries	5,239	male: 47.9 female: 51.5 non-binary: 0.4 PNTS*: 0.2	18-24: 14.4 25-34: 15.7 35-44: 15.7 45-54: 15.2 55+: 39.0	no formal education: 0.2 primary school: 3.0 secondary school: 39.6 bachelor or master degree: 37.6 PhD: 1.8 professional degree: 13.7 other: 3.4 PNTS: 0.6	yes: 9.6 no: 82.3 no work history: 7.1 PNTS: 1.1	yes: 10.2 no: 86.9 PNTS: 3.0	yes: 33.0 no: 65.4 PNTS: 1.6

* Pre-determined quotas used to reflect population characteristics in each country; PNTS: answer option 'prefer not to say'

OVERVIEW OF RESULTS

This section provides an overview of the consultation results for the sample at the time of the report. Since the consultation is still ongoing, at this point the results are presented for the full sample without differentiation for countries or subgroups. A detailed analysis will be presented once the citizen consultation is completed.

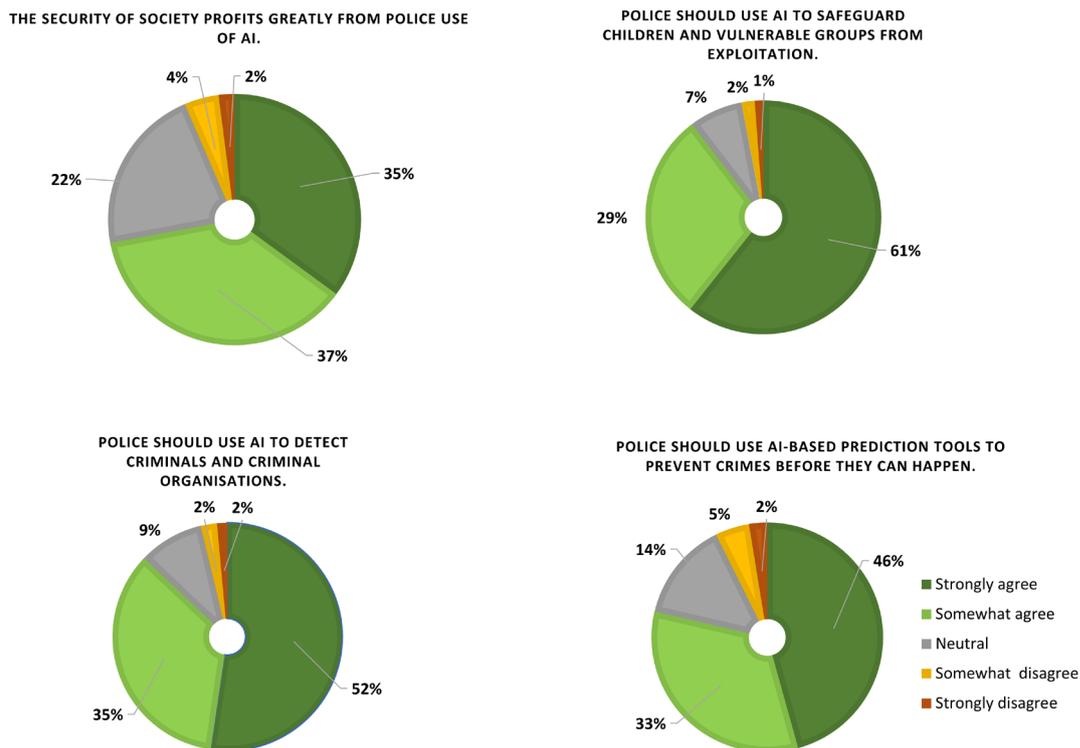
General attitudes towards AI use by police

Participants saw considerable benefits in AI deployments generally with 72.1% agreeing or strongly agreeing that AI can greatly profit society (Figure 1). Even higher was the approval for specific application areas: 89.7% agreed or strongly agreed that AI should be used for the protection of children and vulnerable groups, 87.1% agreed or strongly agreed that AI should be used to detect criminals and criminal organisations and still 78.6% agreed to AI being used to predict crimes before they happen.

This suggests that citizens overall find considerable value in and are in favour of police use of AI if it helps to protect vulnerable groups and society in a meaningful way.

Figure1: Perception of AI use – overall benefit and for specific application areas

BENEFITS OF AI (all countries)

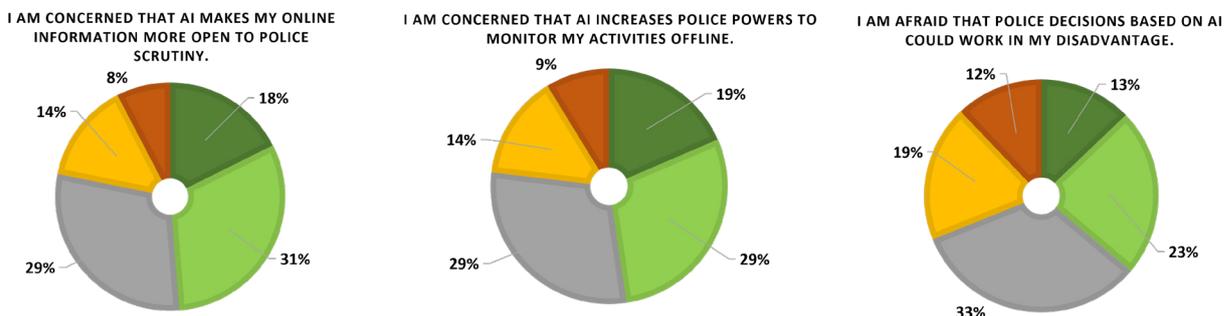


Half of the respondents indicated some or strong concerns about the possibility that AI might make their online information or offline activities more open to police scrutiny (48.7% and 47.7%, respectively). In contrast, a third remained neutral (29.5% and 29.2%, respectively), while about 20% disagreed or strongly disagreed to being concerned (21.8%, 23.1%; cp. Figure 2). Potential negative effects of biased decisions by AI were feared by 36.2%, with the remaining participants equally split between neutral (32.7%) or indicating no fear about potential negative effects (31.1%)

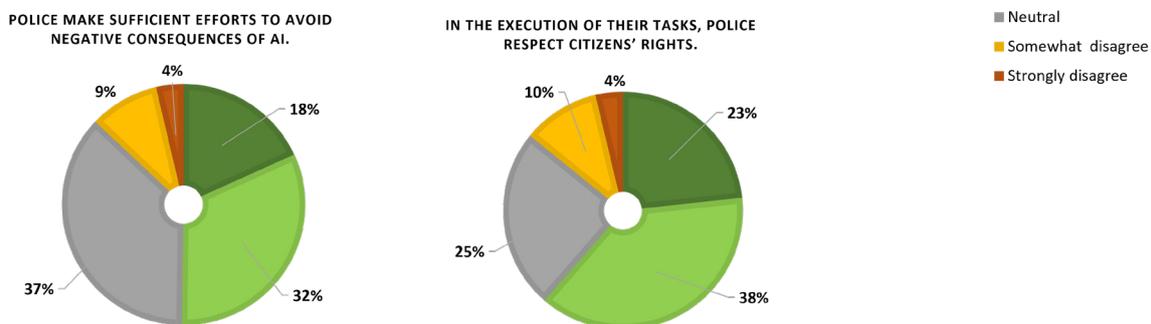
The overall perception of police was positive with 61.4% agreeing or strongly agreeing that the police respect citizens’ rights in the execution of their tasks (versus 14.1% disagreeing or strongly disagreeing). A minority (12.9%) indicated that police did not do enough to avoid negative consequences of AI, while 50.2% agreed or strongly agreed that police make sufficient efforts to avoid negative consequences.

Figure 2: Concerns about AI use and perceptions of police

CONCERNS ABOUT AI (all countries)



PERCEPTIONS OF POLICE (all countries)



Citizen perspective on accountability for AI deployments by police

While citizens were positive towards the potential of AI (see previous section), they also felt strongly that police should be held to account: 92.1% expect police to be held accountable for the way the use AI, 92.1% for the consequences of their AI use (Figure 3). This suggests that citizens expect strong mechanisms, as well as reassurance that police is willing to deploy AI in an appropriate way.

Currently, however, only a third of participants (31%) consider existing mechanisms as appropriate. 26% see them as too weak, while 9% rated them as too restrictive. A considerable number of participants (34%) indicated that they “don’t know” whether current accountability mechanisms are appropriate. The latter implies that a considerable part of the public may lack sufficient information about existing mechanisms to make an informed judgement.

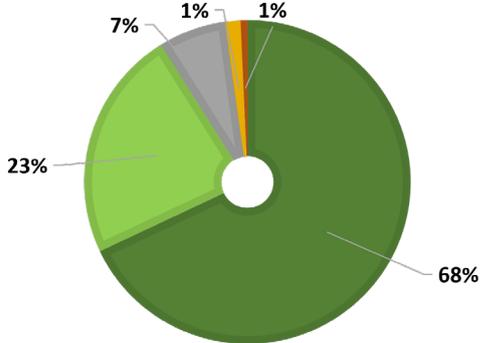
Asked explicitly about the creation of a universal Accountability Framework, the vast majority (82.5%) of participants rated it as either important (29.5%) or extremely important (53.0%) as a way to ensure accountability, compared to only 2.8% who found it of low or no importance.

An overarching AI Accountability Framework thus seems to find broad citizen approval, which gives confidence that the AP4AI Framework will be an accepted approach.

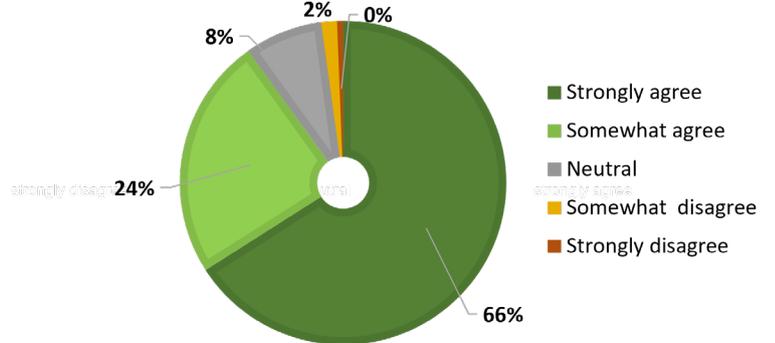
Figure 3: Opinions about AI Accountability

HOLDING POLICE ACCOUNTABLE (all countries)

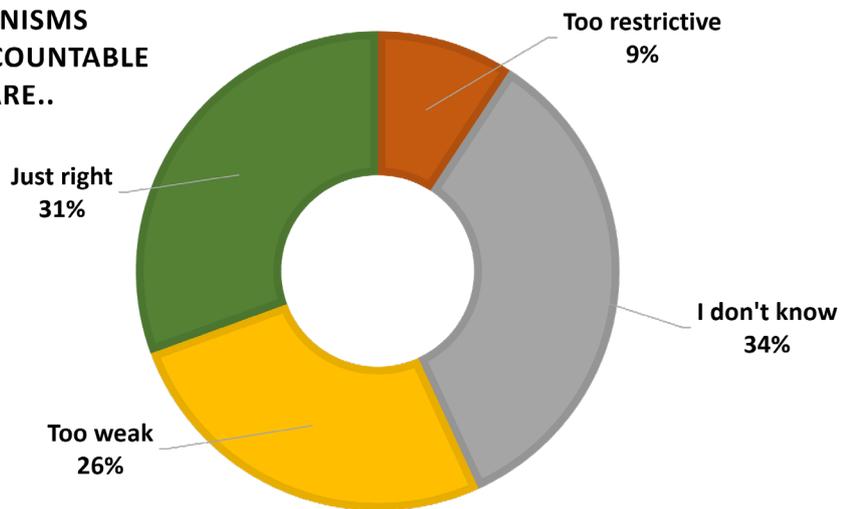
POLICE SHOULD BE HELD FULLY ACCOUNTABLE FOR THE MANNER IN WHICH THEY USE AI.



POLICE SHOULD BE HELD FULLY ACCOUNTABLE FOR THE CONSEQUENCES OF THEIR AI USE.

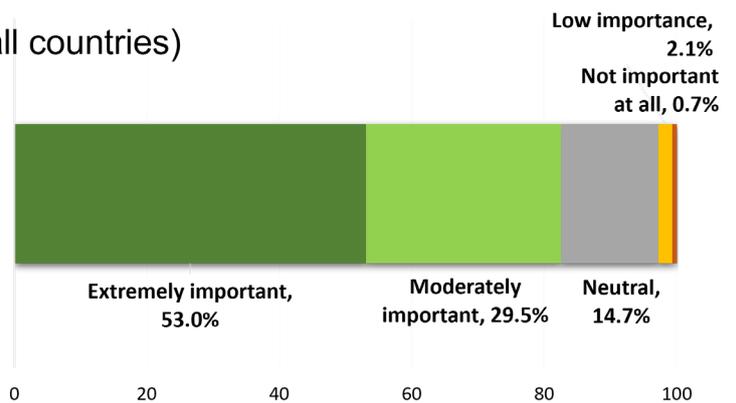


THE CURRENT MECHANISMS FOR HOLDING POLICE ACCOUNTABLE FOR THEIR AI USE ARE..



AI ACCOUNTABILITY FRAMEWORK (all countries)

HOW IMPORTANT IS IT THAT A UNIVERSAL FRAMEWORK IS CREATED THAT ENSURES THE ACCOUNTABILITY OF AI USE BY POLICE.

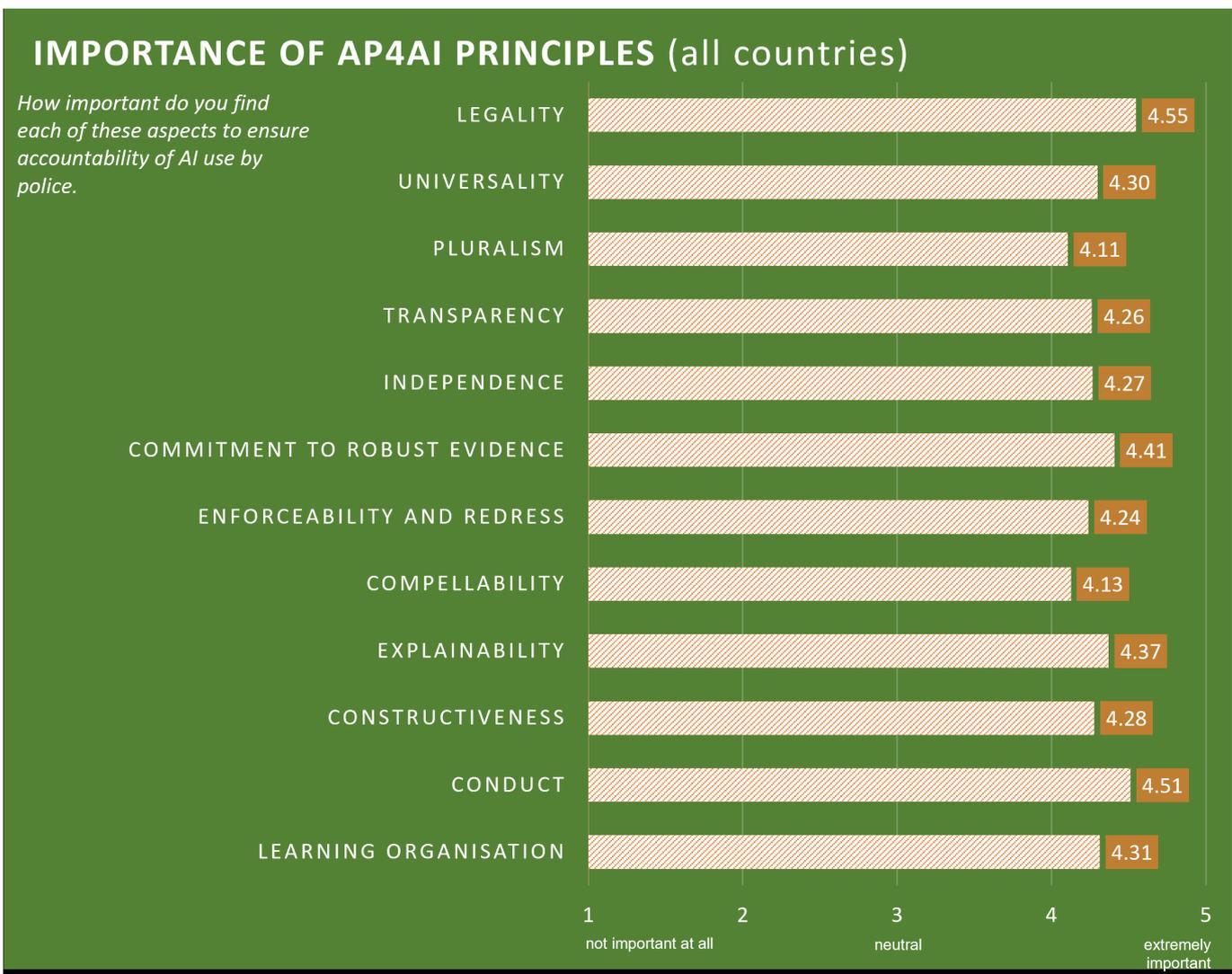


Evaluation of the initial set of AP4AI Principles

Further to a more general understanding of public attitudes towards AI and Accountability, the core purpose of the citizen consultation is to obtain insights into reactions to the 12 AP4AI Principles, as developed in Cycle 1 of the AP4AI Project. It therefore asked explicitly how important participants rated each of the principles from 'not important at all (1)' to 'extremely important (5)'. Figure 4 presents the average ratings across all participants and countries.

As this overview shows, all 12 Principles emerged as important. Legality and Conduct received the highest ratings, Pluralism and Conduct the lowest, although differences amongst the 12 Principles are small. This result validates the relevance of all 12 AP4AI Principles and gives confidence that the 12 Principles are a meaningful foundation for the AI Accountability Framework.

Figure 4: Importance of AP4AI Principles



The consultation also offered the possibility to name additional aspects or principles. Nearly 50% of participants utilised the opportunity providing a total of 2,552 entries. A systematic analysis of these answers will be done once the citizen consultation is completed. However, Table 2 gives a first impression of the richness of this information and the degree of vivid details and refinements to the existing principles they can provide.

Table 2: Examples of the additional information provided in the consultation to the question: “What else should be done to give you confidence that police are using AI in an appropriate way?”

Principle addressed	Examples (presented as written, only typos removed)
Legality	There should be a binding law to protect both parties, the citizens and the government forces that have all access to individual privacy information
Explainability	Just that all actions can be seen and explained clearly with a set reason. No exceptions especially for the rich and the prime minister; To know each step that the police are going through
Transparency	Give up information about AI and what it is and how it’s used and what effects it could have and ask for our opinions; All findings and procedures made aware to the public; Informative videos on social media and the police and govt websites
Commitment to robust evidence	Must be reliable and evidence based for use in a court of law; Just being able to prove any time and any place to show legitimate reason for using AI
Independence	Use of police AI has to be free from political pressure from government/politicians seeking to exploit information gained for their own propaganda; The part where an independent company that’s not police monitors the usage
Conduct	They are abiding by their oath, to serve and protect the public, and to be open for investigation if it is called for; Regular vetting of staff operating or reviewing AI and regular external review of department attitudes to minority groups to assess for bias, racism, misogynistic attitudes that can develop if left unchecked; Employing people of good behaviour and attitude
Learning organisation	Proof that lesson learning translates into positive change; Proper training in appropriate procedures to be used; Constantly have unbiased reviews

Parties responsible for AI Accountability

The expert consultation in Cycle 1 brought up a large number of stakeholders, which experts suggest involving in the AI Accountability process.²⁸¹ The citizen consultation asked the same question to understand the public perception on who should be involved in and responsible for holding internal security practitioners to account.

Citizens show clear preferences for the groups and organisations which should be responsible for (a) the monitoring and (b) the enforcement of corrects and penalties as part of the accountability process. Courts emerged as the preferred body for both areas, followed by police themselves and government/ministries (cp. Figure 5). That police emerged as responsible party – although more for monitoring than for the enforcement of corrections and penalties – is an interesting observation, as it means citizens do trust and expect police to be part of AI Accountability. Interestingly, only a relatively small proportion of participants called on citizens to be part of the accountability process, either in a direct process or through representation. Especially for enforcement, citizens were only considered by 9-10% of participants. As least relevant emerged the inclusion of industry.

On the other hand, nearly 18% of participants prefer to explicitly exclude citizens from monitoring and assessing the police use of AI, similarly to industry and police (Figure 6). Additional groups were mentioned in the open answer option, key amongst them the exclusion of 'governments', 'politicians' and 'criminals'. 40% indicated that no exceptions should be allowed. These answers have implications for the AP4AI Principle of Pluralism in that it provides concrete pointers on how to implement and contextualise stakeholder involvement.

Expert consultations in Cycle 1 further highlighted the need to consider potential exceptions, especially in the application of Accountability Principles such as Transparency and Compellability. The citizen consultation equally made provision for exceptions, mostly in case of time-critical decisions and if information can help criminals to avoid police (Figure 7). Only about 18% refused to permit exceptions, which suggests that citizens are generally sensitive to the complexity of AI Accountability in the internal security domain.

Figure 5: Parties responsible for the Accountability process

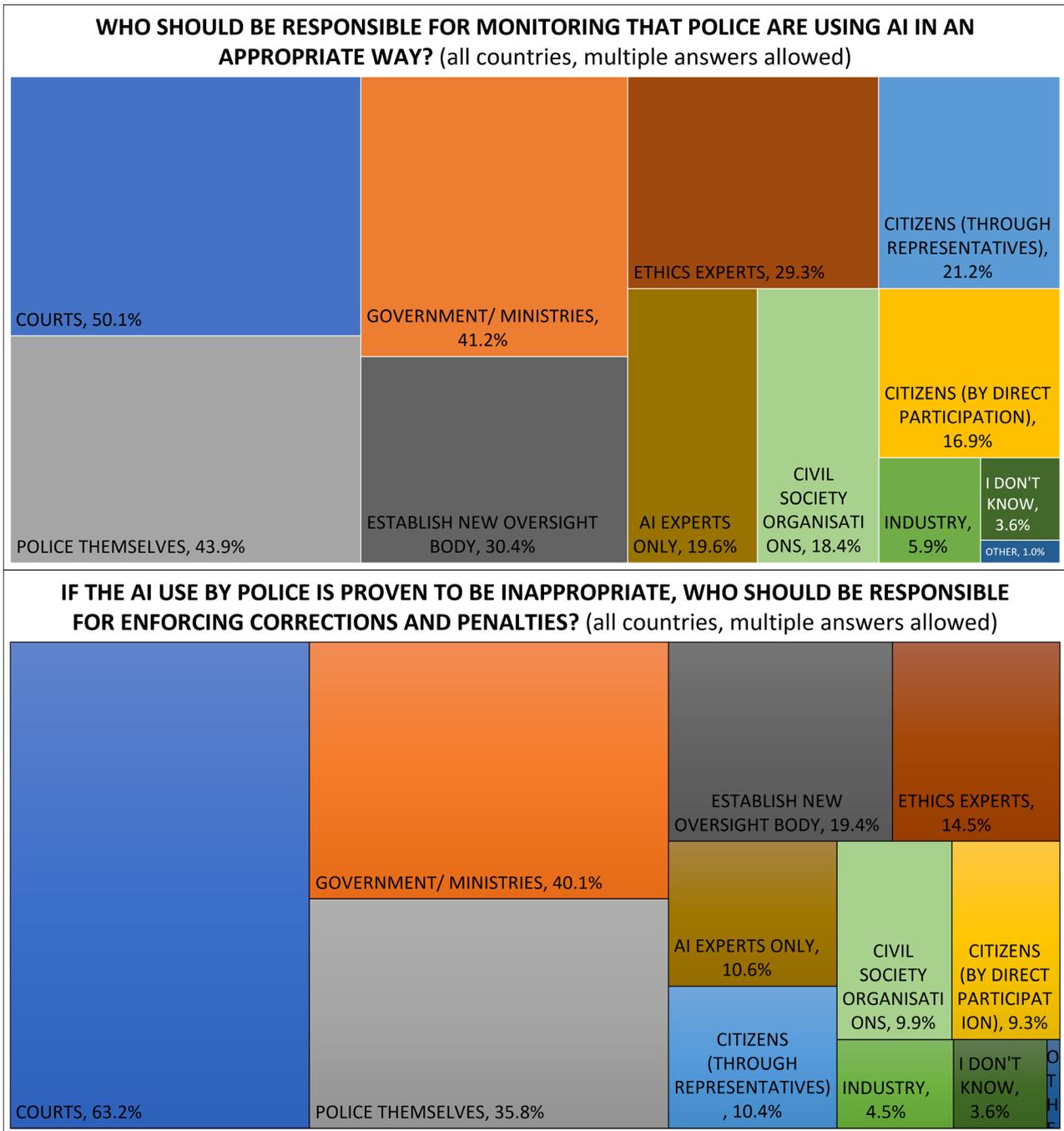


Figure 6: Exclusion of groups from the Accountability process

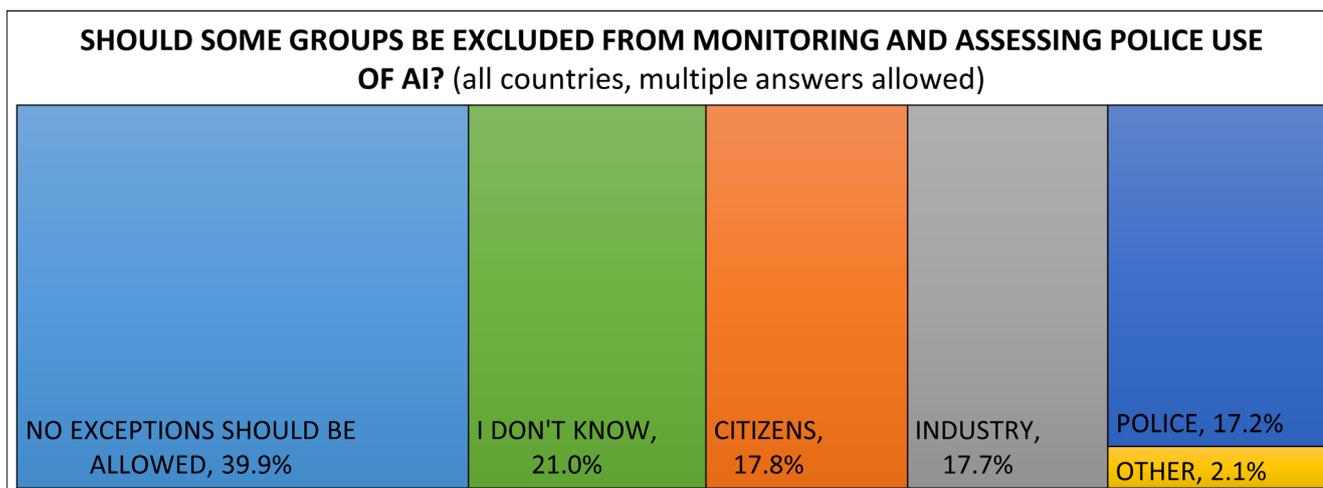
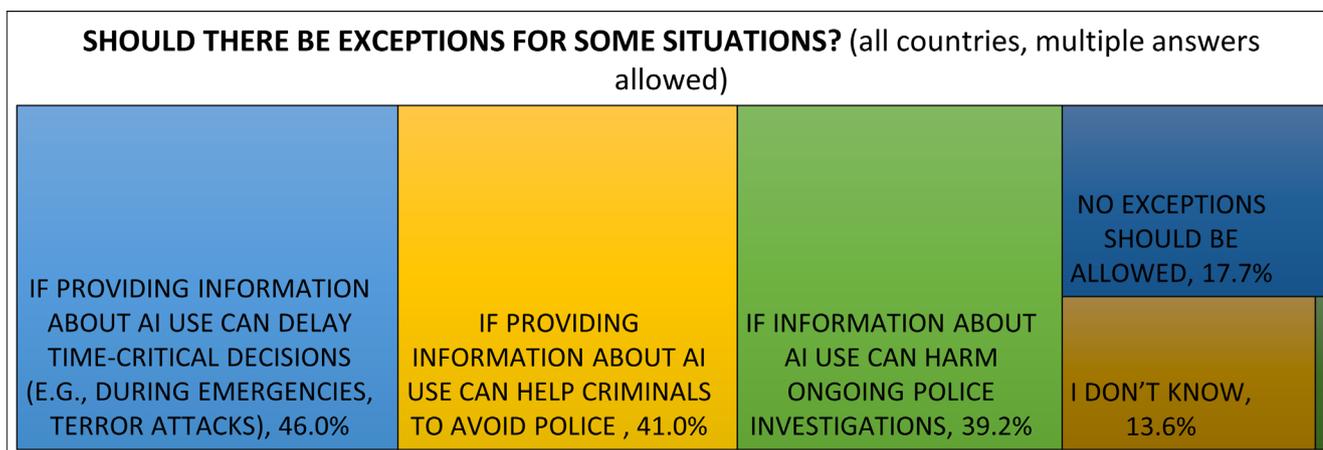


Figure 7: Situations that may warrant exceptions

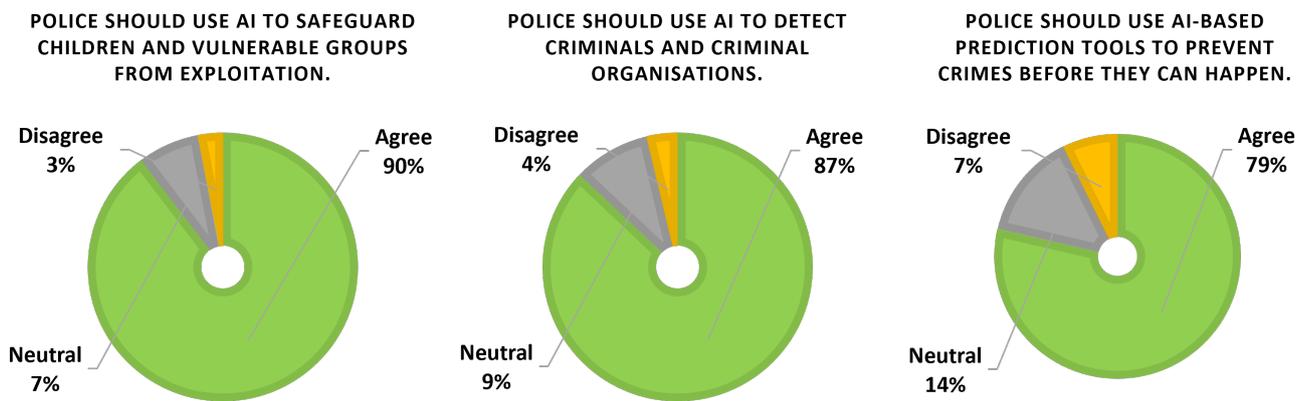


REFLECTION ON FINDINGS FOR AP4AI

The findings presented above offer important insights for the further work of AP4AI and the perception of AI use by internal security practitioners more generally. While they highlight that concerns do exist about the AI use by police forces, they also indicate that citizens seem to see great potential in AI use for safeguarding vulnerable groups and society, including the prevention of future crimes (Figure 8). These observations provide a clear mandate for internal security practitioners that utilisation of accountable use of AI is not only require for their mission and objective in protecting and safeguarding of society but also an expectation from the citizen.

There seems further a strong appetite for Accountability mechanisms, especially given that only a third of participants considered current mechanisms to be adequate. An important result is the importance citizens gave to a universal Accountability Framework, as well as the very high level of importance given to the 12 AP4AI Principles to guarantee AI Accountability in the policing domain.

Figure 8: Benefits of AI deployments by police



The citizen consultation is still ongoing, and the current results are thus only a first impression of the consultation, albeit from a considerable number of citizens and across a highly diverse set of participants. That reactions point largely in the same directions, despite the high diversity of participants, suggests that there is a basis for agreements on the nature of an AI Accountability process for the internal security domain.

More detailed analysis will be conducted once the full dataset is available, especially investigations on whether sub-groups (e.g., in terms of demographics, AI expertise, professional background, self-ascribed minority status, etc.) may differ in their perceptions and expectations and in which way this may influence the contextualisation of an AI Accountability process. In this context, the rich qualitative data we could only hint to in this report will be a vital input for the further development and refinement of the AP4AI Principles, specifically for the creation of Accountability mechanisms that can find broad acceptance within society.

Overall, however, these first results in fact indicate a strong mandate for internal security practitioners to deploy AI. They moreover indicate a strong expectation for the deployment of AI and a strong appreciation of AI Accountability Principles along the lines proposed by AP4AI.

AP4AI FRAMEWORK BLUEPRINT

This section outlines the AP4AI Framework – describing its ambition, foundation in the 12 AP4AI Principles and high-level recommendations for a pathway to the implementation of the Framework in practice. The Framework as outlined below is a first iteration from the synthesis of results in Cycle 1 and Cycle 2 (see [section on AP4AI approach](#)).

Framework in AP4AI is defined “a high level semantically coherent and objective driven conceptual model which includes a specific set of concepts, processes, procedures, tools and methodology(ies) in support of a particular thinking paradigm.”²⁸²

The AP4AI Framework defines its practice-oriented nonlinear conceptual model based on previous research and practices on policing, accountability, AI, and AI accountability, the expert consultations in Cycle 1 (see [AP4AI Summary Report on Expert Consultations](#))²⁸³, gap analysis and critical review of related legislations, directives, policies, practices and academic research, as well as the citizen consultation in Cycle 2 (see [section on Citizen consultation](#)).

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 AI Accountability when researching , designing, (de) commissioning, procuring and utilising AI and (b) oversight bodies and the public to measure security practitioners’ use of AI against.²⁸⁴

In pursuit of the above ambition the AP4AI Framework consists of two core elements:

- the 12 Accountability Principles and
- application guidelines for their implementation into operational environments

The 12 Accountability Principles (see Table 3) define the requirements that need to be fulfilled to assure Accountability for AI utilisation in the internal security domain. The 12 Principles are the foundation on which all other AP4AI activities and solutions are built. The *AP4AI Report on Expert Consultations*²⁸⁵ presents the Principles in a uniform structure (definition, practical considerations, examples of applicable laws, and where applicable, elaborations and examples). In the current report they are elaborated towards an initial method for implementation.

Table 3: List of AP4AI Principles

<p>Legality: Legality means that all aspects of the use of AI should be lawful and governed by formal, promulgated rules. It 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.</p>	<p>Enforceability and Redress: Enforceability and redress requires mechanisms to be established that facilitate independent and effective oversight in respect of the use of AI in the internal security community, as well as mechanisms to respond appropriately to instances of non-compliance with applicable obligations by those deploying AI in a criminal justice context.</p>
<p>Universality: Universality provides that <i>all</i> relevant aspects of AI deployments within the internal security community are covered through the accountability process. This includes all processes, including design, development and supply, domains, aspects of police mission, AI systems, stages in the AI lifecycle or usage purposes.</p>	<p>Compellability: Compellability 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.</p>
<p>Pluralism: Pluralism ensures that oversight involves all relevant stakeholders engaged in and affected by a specific AI deployment. Pluralism avoids homogeneity and thus a tendency or perception for the regulators to take a one-sided approach.</p>	<p>Explainability: Explainability requires those using AI to ensure that information about this use is provided in a meaningful way that is accessible and easily understood by the relevant participants/audiences.</p>
<p>Transparency: Transparency involves making available clear, accurate and meaningful information about AI processes and specific deployment pertinent for assessing and enforcing accountability. 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.</p>	<p>Constructiveness: Constructiveness 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.</p>
<p>Independence: Independence refers to the status of competent authorities performing oversight functions in respect of achieving accountability. This applies in a personal, political, financial and functional way, with no conflict of interest in any sense.</p>	<p>Conduct: Conduct governs 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.</p>
<p>Commitment to Robust evidence: 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.</p>	<p>Learning Organisation: Learning Organisation promotes the willingness and ability of organisations and people to improve AI 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 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.</p>

From the outset, the AP4AI Project aimed at translating the Accountability Principles (as conceptual representation of AI Accountability requirements) into actionable steps and processes in support of internal security practitioners). In this report, therefore, each of the principles has been qualified with a contextualisation for concrete AI deployment within the internal security domain, providing legal and practical consideration, as well as examples. This translation for practical application is the second core element of the AP4AI Framework. The tangible realisation of the Principles is demonstrated through provision of an implementation container which will serve as a mechanism for the implementation of the principles while providing concrete accountability narratives. It will provide flexibility for local implementation at the organisational level.

The outline of an implementation mechanism is a first step in creating a practice-oriented Framework. Refinements and contextualisation will continue along with the ongoing expert consultations (Project Cycle 3, see [section on AP4AI approach](#)).

OUTLINE OF MECHANISMS FOR THE PRACTICAL APPLICATION OF AP4AI

One of the strengths of the AP4AI Framework is its ambition for practical application. The practical application of the AP4AI Principles to a 'live' problem will however contain many variables and knowledge of their interdependencies, not all of which will be apparent from the outset nor to a single stakeholder. AP4AI advocates for an **AI Accountability Agreement (AAA)** that identifies the relevant accountability provisions for each application of AI. While not a legal document or enforceable contract, the AAA commits parties to the approach that each will take towards a formal and implementable processes for the application of the *Accountability Principles* for different uses of AI within the internal security domain.

An **AI Accountability Agreement (AAA)** should be viewed as a **social contract underpinned by legal obligations** between internal security organisations and its stakeholders including citizens, oversight bodies, suppliers, consumers of AI services (e.g., other agencies) and others, as applicable. The AAA can thus be understood as an implementation container or reference architecture,²⁸⁶ which drives implementation of the Principles in a practical and operational settings of internal security organisations. It hence serves as a mechanism to bring the abstract nature of the principles into the implementable environment of internal security organisations and their wider ecosystem (e.g., oversight bodies and government agencies).

The AAA must be created and validated prior to any programme of work that encompasses the application of AI. Each application of AI involves one or more stages of the AI lifecycle: scoping planning, research, design, development, procurement, customisation, deployment, modification maintenance and decommissioning. Each stage of the AI lifecycle may require a new or updated AAA that balances the critical and non-negotiable elements surrounding the application (as introduced by the Accountability Principles, such as adhering to strict legal obligations or a guarantee of professional conduct in all aspects of AI use), while providing flexibility for the use of the application under operational discretion.

For example, the AAA may recognise the necessity of operational decision makers' discretion, within certain boundaries, to change some aspects of an operation to use Live Facial Recognition (LFR) for practical operational reasons such as specific times and places (cp. also exceptions such as discussed in the expert consultations in Cycle 1²⁸⁷ or as reflected in the citizen consultation above). However, the AAA should also account for the scenario where such changes, which might ordinarily be uncontroversial, will involve greater sensitivity – and therefore greater accountability considerations. In the context of LFR this may refer to revised geographic areas and neighbourhoods, particular sites (places of education and worship) or events (political rallies, demonstrations and elections). Therefore, although an internal variation may appear to be unexceptional and permitted from a strictly project management perspective, such a change in the given example would represent a critical variation in terms of *accountability* and should therefore be anticipated and encapsulated within the AAA.

In the review of existing AI frameworks and regulations earlier in this report and the *AP4AI Report on Expert Consultations*²⁸⁸ we have discussed and elaborated on the notion of 'AI impact'. In order to pave the way for the implementation of the 12 Accountability Principles, AP4AI utilises the concept of Materiality.²⁸⁹

Materiality is an assessment of the relative impact that something may have on accountability within the context of an application of AI in the internal security ecosystem. A **materiality threshold** is an important component of the AAA, as can be seen in the above example, where the material importance and impact of a change of date or location will very much depend on the nature of the AI project.

It is important for each application of AI that internal security practitioners (e.g., LEAs) identify and assess the materiality of each Principle **in terms of accountability** and record them in the AAA. Legality and Transparency are good examples. The strict legal obligations for internal security practitioners will generally allow little discretion when exercising their functions at organisational level and *all* legal requirements will be sufficiently 'material' to their accountability. The key consideration for materiality in the context of accountability is that the issue has been expressly *considered and determined* rather than having been overlooked or later explained away as irrelevant or inconsequential.

Considering the example of automated decision making (ADM), perceptually, there is a material difference between the police use of automated decision making to manage the replacement of uniforms and their use of automated decision making to issue a penalty notice to a citizen. Applications of AI that involve purely 'administrative' elements not connected with a 'policing purpose' are still closely controlled by data protection legislation but do not require the same level of accountability as those employed to assist in upholding the law. Bringing the above together, in a practical sense, the AP4AI Principles can firstly help to identify:

1. the accountability *must-haves* (non-negotiables), *should-haves* and *could-haves*²⁹⁰ within the specific application of AI
2. who *will be* Responsible, Consulted and Informed (RACI index²⁹¹) in relation to each of the AP4AI Principles for each application of AI and who *has been*

Consulted and Informed about the purpose and development of the AI application (with a summary of what they have said)

3. the materiality *thresholds and tolerances* to allow for practical variance (dates, changes in personnel, etc.), the range of acceptability and for assessing the proportionality of disclosure, consultation, and publishing of information
4. the process that must be followed before making any variation to the specific application of AI

The AAA should clearly set out and **formalise** these four steps, it should be signed at an executive level and **published** as a formal decision to the relevant bodies. Thus, the AAA will be a conspicuous identification of and commitment and social contract to the accountability provisions that will apply to the AI project from the outset. Any future variance to the AAA must be clearly documented and all previous versions should be retained in their original form. Parties to the AAA may be one internal security stakeholder (e.g., LEA) but also actors in the research and development or suppliers of various AI applications for the internal security domain.

The AAA should address all AP4AI Principles and their realisation in an operational setting for the specific application of AI (see Figure 9). To achieve this, the AAA must include, as a baseline, the following four phases as components: **context, scope, methodology, and accountability governance**. Each phase of the AAA should adopt the application of the 12 Principles and use them as a milestone to progress to the next stage.²⁹²

Figure 9: Stages of development for an AI Accountability Agreement

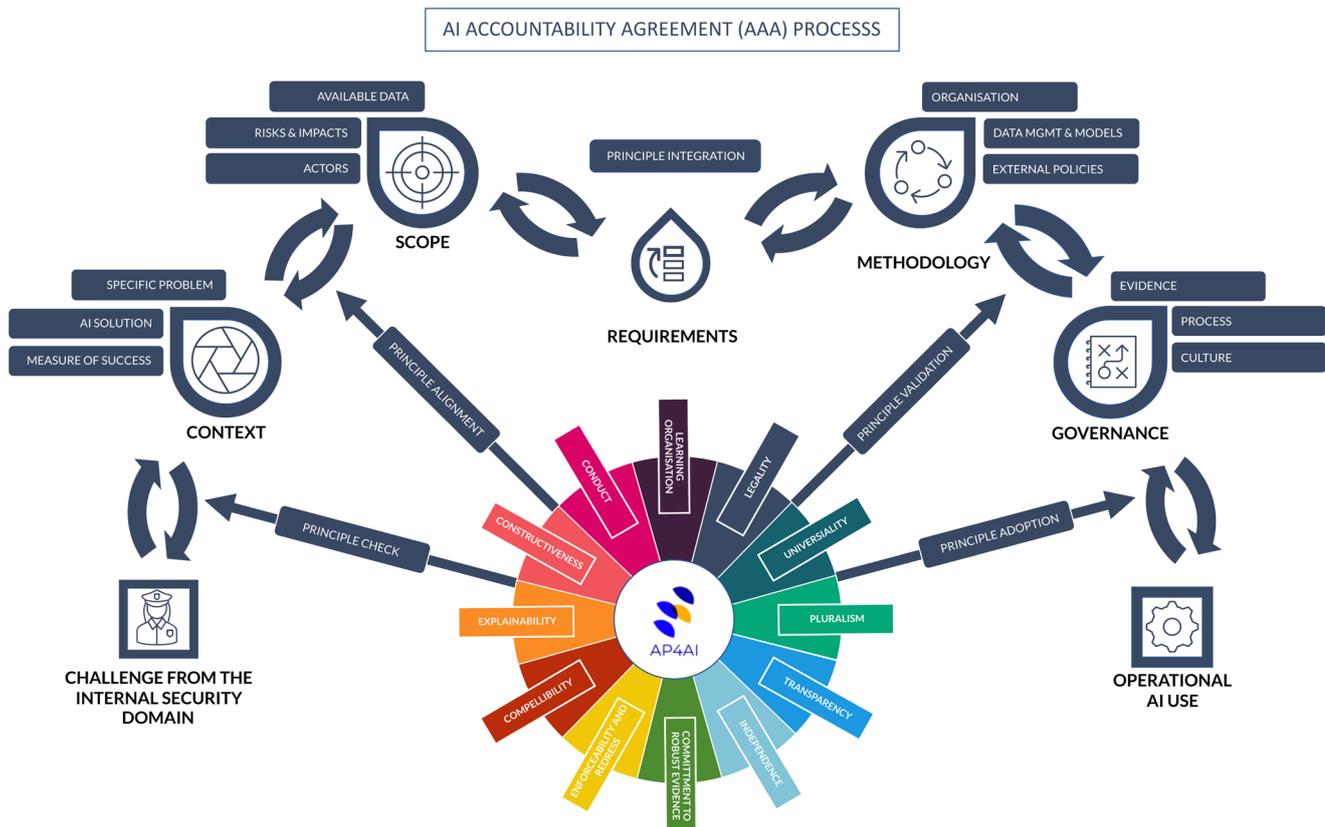


Figure 9 outlines and demonstrates how conceptually the AAA will be initiated and evolve against the Accountability Principles, from the identification of a specific challenge in the internal security domain (e.g., utilisation of AI to combat child sexual exploitation or usage of AI-driven big data analytics for the prevention of a terrorist attack) through the elaboration of the context and scope into the requirements, the implementation methodology and finally the governance mechanisms for accountability to enable the AI application to enter operational use. It should be noted that AP4AI only advocates the four components of the AAA, namely **context**, **scope**, **methodology**, and **accountability governance**. The particular instantiation or organisation-specific realisation based on their own AI development life cycle²⁹³ can be adapted and can differ from the detailed view of the AAA as proposed here.

In the following section, we provide an example of an AAA as a prognosis solution rather than a diagnostic description for implementation.

AAA: Context

The context encompasses the entire application of AI and the resulting capabilities for its use in the internal security domain. Embracing research, design, development, testing, procurement, deployment and modification efforts, as well as the need for long-term monitoring, the context must present the rationale for the uptake of an AI application; specifically, the lawful purpose which the application of AI will support and the necessity for utilising an AI application over other available methods. This should be followed by explanations on how the application of AI can further the organisation in solving the specific problem and what assumptions (if any) have been made. A clear indication of the value proposition, the scope and expected impact of the application of AI on precise business processes and the anticipated benefit(s) should also be documented. Finally, as a measurement of success, defining the critical success factors (CSFs) and key performance indicators (KPIs) should provide a realistic metric for evaluating the application alongside any potential risks or foreseen negative consequences. It further needs to afford the identification and assessment of emerging and unexpected consequences. In essence, the context must provide a detailed view on the vision and intention for the application of AI and how it is aligned to the strategic goals of the organisation and the specific problem it aims to help resolve.

AAA: Scope

The scope element of AAA provides the next level of detail and boundaries in the accountability assessment. It should follow an initial threshold assessment against all AP4AI Principles and provide more definitive details than available during the context phase. The scope must begin by restating the exact purpose of the application of AI by practitioners. It should apply the concept of purpose limitation and carry out a data protection impact assessment. This should lead to explicit inclusion and exclusion criteria on *what*, *how*, *when* and *to who* remaining mindful of the boundaries and tolerances discussed above. A human rights impact assessment (HRIA) together with a specific AI-risk impact assessment, stakeholder

engagement and identification of required consultation processes should result in the characterisation of an initial set of requirements for the application of AI. The requirements should plainly state the data requirements, type (e.g., rule-based/statistical, supervised/unsupervised, use of deep learning, etc.) and exact functionality of the application of AI.

At this stage it should also be possible to identify individuals and groups that are affected directly and indirectly and provide an assessment of possible sources of implicit and explicit bias (cp. Pluralism Principle). Organisationally, the scope should align with an organisation's current mission and overarching corporate strategy. Further, consultations with oversight bodies on how monitoring processes will be managed should begin. Expectations for project management processes and financial targets/impact also need to be documented at this stage if they are applicable.

AAA: Methodology

The methodology component can apply to some or all operational aspects, namely development, procurement, implementation, deployment and modification. The methodology phase captures the requirements and considerations for the actual utilisation of the system. The designated project team and key contacts run alongside an accurate and auditable RACI chart²⁹⁴ (for both internal and external parties). In this phase exact protocols for engaging in co-creation sessions and technical operations should be made explicit. Any underlying assumptions in scope or modelling, semantics or limitations on interpretation must also be addressed or documented.

For development (initial, maintenance and updating), the approach to the software development lifecycle should be defined and lines of communication established. Moreover, replicable processes for data modelling, training, testing, refinement, evaluation and updates are required along with mechanisms for continuous monitoring.²⁹⁵ As the application of AI is coming close to operational use, guidance through specific use cases for AI deployment scenarios, the actors, boundaries and thresholds should be laid down and evaluation check points and targets set. All organisational policies considering mechanisms and processes for redress, engagement with oversight bodies and standardisation of evidential capture must also be in place.

AAA: Accountability Governance

The accountability governance mechanisms are enacted when the application of AI is operating as a live system. This must include records for how accountability will be governed ensuring appropriate documentary evidence and methods for monitoring and evaluation of accountability as well as procedures for sharing this information where necessary (cp. Principles of Transparency, Explainability).

The documentation must include the selection and inclusion of responsible organisations for independent evaluation and the method and timeline of their engagement (cp. Independence Principle). The approach also calls for embedding

accountability within any application of AI from development (accountability-by-design), and provision of training and awareness to ensure a cultural shift that adopts accountability as a fundamental enabler of exploitation of AI by internal security practitioners (cp. Learning Organisation Principle).

To develop and realise the AAA in an operational setting, the stakeholders, bodies and parties to the agreement must understand what it means to apply each principle in practice. In the following section, we present these considerations, the potential thresholds and application of materiality and the business processes and functional constraints that each organisation must embed into their practices to assure their use and adoption of AI is in line with each of the 12 Accountability Principles and, ultimately, the AAA.

The next section provides the foundation of a practical guide for organisational implementation of each Accountability Principle for the application of AI in the internal security domain.

AP4AI ACCOUNTABILITY PRINCIPLES

This section provides a structured, semantic representation for each Accountability Principle as part of the initial implementation guidance. The template used to present each principle consists of eight elements which collectively provide the core requirements for the systematic implementation of the AP4AI Principles for research, design, development, procurement, deployment and modification of AI in the internal security domain. The template is designed in a way that it can be extended and refined throughout the AP4AI Project, yet maintain its conceptual foundation which is grounded in evidence-based research, as well as input from expert and citizen consultations. It builds on and expands the original definitions provided in the [AP4AI Summary Report on Expert Consultations](#).²⁹⁶ The granularity (e.g., set of purposeful questions) and visual representation of the 'implementation guide' in each principle will support the development of a software tool in future stages of the AP4AI Project.

Guide to the Principle presentation

Name – Principle name, validated in expert consultations

Meaning – provides the Principle definition contextualised for AI and the internal security domain

Materiality threshold – offers an assessment of the relative impact that something may have on accountability within AI development or utilisation

Examples of applicable law – lists examples of applicable law pertinent to AI Accountability in the internal security domain

Note on Human Right Impact Assessment – provides an initial direction for HRIAs and alerts the reader about the pivotal role of HRIAs in the context of AI Accountability Principles

Note on Data Protection Impact Assessment (DPIA, where applicable) – alerts the reader to legal and ethical requirements of conducting a DPIA and, where applicable, a Privacy Impact Assessment (PIA)

Implementation guide – identifies the processes, activities, tasks, documentations, assessments, actions and communication needed for the realisation of the Principle

Operational considerations – provides clarification and further consideration about implementation of the principles for the operational environment

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.

Materiality threshold

The legal obligations for internal security practitioners will generally allow little discretion when exercising their functions at organisational level and *all* legal requirements will be material to their accountability. However, AI-led systems used by internal security practitioners may involve purely 'administrative' elements that are not connected with a 'policing purpose'. An example would be the use of automated decision-making which is closely controlled by data protection legislation generally but there is a material difference between the use of automated decision-making to manage the replacement of uniform and automated decision-making to issue a penalty notice to a citizen. Similarly, all policing activity is to some extent intrusive, and *operational* discretion can often be involved in policing whereby the decision-maker on the ground is given a degree of latitude without the issue becoming unlawful itself. It is important for the AI programme to identify areas where this is to be left to the operational commander and to assess the materiality **in terms of accountability**. Compliance with all relevant legal duties will be material for the purposes of this principle and the only practicable scope for an assessment of 'materiality' will be in the context of the general legal '*de minimis*' maxim ("the law is not concerned with trivialities"). Examples might include very minor disagreement over the literal wording in a contractual provision which has no bearing on the performance of the contract or on the other Accountability Principles.

Examples of applicable laws²⁹⁷

- 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.²⁹⁸
- 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.

Note on Human Right Impact Assessment

Human Rights provide a general and largely universal framework for the principle of Legality, in terms of safeguarding individual and social rights and freedoms. Given the broad impact of AI applications many human rights and freedoms are potentially impacted. Although Human Right Impact Assessment (HRIA) practices have recently been elaborated in this field,²⁹⁹ there is still limited development of them in the field of AI. While traditional HRIA methodologies and models are designed for large-scale territorial projects and to provide policy guidelines, HRIA in the AI context requires a different approach focusing mainly on **prior assessment**, a **human rights-by design approach** and a **formal risk evaluation** based on the likelihood and severity of potential impacts.³⁰⁰ Further general guidelines that are also applicable to the AI sector are the UN Guiding Principles on Business and Human Rights,³⁰¹ and specifically Section II on corporate responsibility to respect Human Rights, which enshrines several key HRIA requirements (stakeholder consultation, regular assessment, transparency, role of experts, etc.). References to the human rights framework are also present in guidelines focusing on ethics issued by private and public bodies,³⁰² although the blurred overlap between ethical principles and Human Rights may make concrete implementation more difficult.

Note on Data Protection Impact Assessment

Data Protection Impact Assessment (DPIA), as well as the older and broader Privacy Impact Assessment (PIA), are more sectoral human rights assessment tools, centred on data protection and privacy. Given the wide adoption of data protection laws worldwide, DPIA and PIA are well-known and widely used instrument to assess the lawfulness of the use of personal data, including its processing in the context of AI. However, the presence of national data protection legislations and national data protection authorities has generated a variety of models and tools for DPIA, also to address the specific nature of the context of their application.³⁰³ In all these models, the lawfulness of personal data processing is a key element and is assessed in relation to the applicable law.

Implementation guide

<p>APPLICATION OF LAW</p> <p>How do the applicable laws apply in the this context?</p>  <p>Check the list of applicable laws</p> <p>Check for infringements against rights and freedoms</p> <p>Involve data protection and human rights experts</p>	<p>QUALITY ASSURANCE</p> <p>What quality assurance and bias mitigation processes do you have in place for the data lifecycle - for both acquired and collected data?</p>  <p>Is it a decision about a legal duty or right of an individual?</p> <p>Reliance on models of human behaviour or characteristics?</p> <p>Methodology to define measures and protocols</p> <p>Methodology to mitigate against harms and risks</p>	<p>PRIVACY HARM</p> <p>Does the use of AI deal with special categories of personal data, as defined by applicable legal norms?</p>  <p>Verify use of special categories of personal data</p> <p>Were there less intrusive alternatives & why is an AI application preferred?</p>
<p>NECESSITY AND PROPORTIONALITY</p> <p>Are the overriding principles of necessity and proportionality complied with?</p>  <p>Is the use of an AI system necessary and proportionate?</p>	<p>DATA PROVIDER AND PURPOSE</p> <p>Will any data being used in the production of the AI system be acquired from a vendor or re-purposed from existing datasets?</p>  <p>Describe the data source and provider organisation</p> <p>Describe the data collection and collation process</p>	<p>OBJECTIVE OVERSIGHT</p> <p>Is the appropriate oversight body engaged, in respect of the activity?</p>  <p>Can an individual prove the wrongness of a decision?</p> <p>If it is about them, can they do so without going to court?</p> <p>Is the harm of a wrong decision fully reversible?</p>
<p>LEGISLATIVE GAPS</p> <p>Some aspects of AI usage, including new developments and capabilities, may not be regulated in existing laws and standards.</p>  <p>Is there case law where no decision was reached?</p>	<p>RESIDUAL RISKS</p> <p>Despite legal compliance, any residual risks particular to AI should be addressed.</p>  <p>Is the AI system designed to be adaptive?</p> <p>Will outputs change due to parameter updates?</p>	<p>EQUALITY</p> <p>An equality impact assessment (EIA) should be conducted considering impacts on affected individuals</p>  <p>Conduct an EIA and cover all relevant protected characteristics</p>
<p>DEMONSTRATION OF COMPLIANCE</p> <p>How can compliance be demonstrated?</p>  <p>Conduct a fundamental rights impact assessment</p> <p>Conduct a legal impact assessment</p>	<p>EXEMPTIONS AND SAFEGUARDS</p> <p>Do any legal exemptions apply? If so, are appropriate safeguards in place?</p>  <p>Can individuals avoid the AI use or decision?</p> <p>Can decisions be taken via a different procedure?</p> <p>Is there an alternative to the technical system?</p>	<p>PUBLIC CONCERNS</p> <p>Legal compliance alone may not address wider public concerns.</p> 

LEGALITY

All aspects and activities of AI accountability must be exercised i

Operational considerations

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. 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.

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 offshoring by the relevant accountable organisation.

Materiality threshold

While all organisations and individuals having a significant impact on/involvement with the AI programme must be subject to the Principles, there will be those whose role is too remote from the inputs/outcomes to be included. Examples might include some people who are purely involved in the technical installation of agreed equipment or provide generic project management support (they can be identified in the project's documentation). Universality applies a holistic, catch-all provision to ensure there are no significant accountability gaps, but there will be many potential impacts and outcomes of the project not all of which will be of sufficient relevance/importance to be included. Similarly, some technical processes may not be of sufficient relevance to accountability to be included.

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.
- 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.

Notes on Human Right Impact Assessment and Data Protection Impact Assessment

In both in HRIA and in DPIA, the assessment covers the entire life cycle of any product/service. In the field of data protection, a key operating principle concerns

the so-called by-design approach, which means that data protection issues must be considered and analysed from the earliest stage of product/service design to mitigate any negative impact on data subjects.³⁰⁴ The same approach is now suggested in literature and in the AIA proposal for AI, adopting a prior assessment of AI applications before their deployment and use.

Implementation guide

UNIVERSALITY

Universality requires that all aspects of AI use fall under the remit of accountability.

<p>AI LIFECYCLE MANAGEMENT</p> <p>Apply to all components and the complete AI system lifecycle, from design to decommissioning.</p>  <p>What is the goal of the AI system?</p> <p>Who is involved in the development process & what are their roles?</p> <p>Who decides which metrics are optimised?</p> <p>Who determines the training/testing datasets?</p> <p>Who determines the feature selection?</p> <p>Who determines error trade-offs and error discrepancies?</p> <p>Map the entire process, roles and feedback loops & communication</p>	<p>OUTCOMES AND IMPACTS</p> <p>Have all outcomes and possible impacts of AI deployment been considered?</p>  <p>Conduct a fundamental rights and legal impact assessment</p> <p>Identify risks, appropriate mitigation measures & safeguards</p> <p>Respect privacy and data protection</p> <p>Cybersecurity and privacy preserving measures</p> <p>Data about marginalised or vulnerable groups</p>	<p>STAKEHOLDERS</p> <p>Have all relevant stakeholders been considered including national regulators and oversight bodies?</p>  <p>Are external stakeholders involved in development?</p> <p>AI ethicists, human rights experts and affected groups involved</p> <p>Is there an alternative to the technical system?</p>
<p>UP / DOWNSTREAM PROCESSES</p> <p>Have all processes affected by AI been accounted for?</p>  <p>Process documentation access and storage location</p>	<p>MEASURING COMPLIANCE</p> <p>How is compliance with this principle measured? Who is responsible for this?</p>  <p>Is there an owner of compliance assessment?</p>	<p>SOCIETY AND INCLUSIVITY</p> <p>What efforts have been made to understand and address concerns and legitimate expectations of specific sections of society and individuals having characteristics requiring additional consideration</p>  <p>How and when are societal actors involved & insights considered?</p>
<p>RESPONSIBILITIES</p> <p>Does everyone understand their responsibilities in respect of compliance with accountability? How is this ensured?</p>  <p>Define and map the responsibilities for each of role</p> <p>What are the developers and users' responsibilities for impact</p> <p>What is the chain of accountability from design to deployment</p> <p>Define the logging protocol for documenting workflow</p> <p>Define the external auditing and oversight for workflow</p>	<p>OVERSIGHT OBLIGATIONS</p> <p>What quality assurance and bias mitigation processes do you have in place for the data lifecycle - for both acquired and collected data?</p>  <p>Who is responsible for system design?</p> <p>Who is responsible for system implementation and outcome?</p> <p>Who is managing end-user feedback and response?</p> <p>Who will respond to doubts and challenges from individuals?</p>	<p>REASONABLE RISK</p> <p>What are the remaining security and privacy risks and why are they reasonable?</p>  <p>Unresolved biases and sources of unfairness</p> <p>What is the potential harm if the AI system is misused?</p>

Operational considerations

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 AI Accountability, any restrictions should be recorded in a specific and clear way, including justifications and mitigating measures adopted in respect of achieving accountability.

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.³⁰⁵

Materiality threshold

The nature of the AI programme (security, confidentiality, data sharing restrictions, etc.) may necessarily mean that only a few organisations are able to be involved. Where this is the case, the fact that Pluralism cannot be achieved as broadly as the Principles would ordinarily encourage should be recognised and recorded (including justification and rationales with appropriate reference to related legislation).

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.

Note on Human Right Impact Assessment

Stakeholder involvement is a key methodological requirement of HRIA. However, the active stakeholder participation can be enabled using a wide range of different techniques, depending on the context and target population.³⁰⁶ Regarding Human Rights, participation can also provide a better understanding of potentially affected rights, including by disaggregating HRIA to focus on specific impacted categories, and a way of taking into account the vernacularisation of Human Rights. Moreover, where AI systems are used in decision-making processes, participation can also be seen as a significant Human Right in itself, namely the right to participate in public affairs.

Note on Data Protection Impact Assessment

Pluralism is also a relevant component in DPIA practice, both in terms of variety of experts involved in carrying out the assessment and in terms of participation of data subjects and relevant stakeholders. In this respect, the GDPR does not require a mandatory engagement of stakeholders but requires the involvement of rightsholders when appropriate.³⁰⁷ However, the same requirement is not present in the Law Enforcement Directive.³⁰⁸

Implementation guide

<p>STAKEHOLDER COVERAGE</p> <p>Is the selection of affected stakeholders sufficiently comprehensive?</p>  <p>Document the methodology for selecting stakeholders</p> <p>Do stakeholders cover local, national, cross-national geography?</p> <p>Explain the absence of any key groups that are missing</p> <p>How are potentially affected stakeholders identified?</p> <p>How is stakeholder coverage documented and by who?</p> <p>Which stakeholders have protected characteristics? What are they?</p> <p>Could this AI system present concerns to specific groups?</p> <p>Which characteristics align with these concerns?</p>	<p>PROCEDURAL INFORMATION</p> <p>Has full information about procedures been provided in a clear and meaningful way, which also achieves the management of expectations?</p>  <p>What information is shared with engaged stakeholders?</p> <p>How are outputs of the system explained to users?</p> <p>How are outputs explained to affected individuals?</p> <p>How is understanding of the outputs verified?</p>	<p>ENGAGEMENT STRATEGIES</p> <p>Have a variety of methods of engagement been implemented, in the true spirit of inclusiveness?</p>  <p>Are there adequate resources for meaningful & diverse engagement?</p>	<p>PLURALISM</p> <p>Ensures participation by all key public and private stakeholders promoting their democratic and collaborative engagement.</p>
<p>STAKEHOLDER EVOLUTION</p> <p>Should stakeholders remain the same at all stages of accountability procedures and engagements?</p>  <p>Document changes to stakeholder groups over time</p>	<p>STAKEHOLDER ROLES</p> <p>Do participants understand their role within the process and the purpose of it?</p>  <p>Who is tasked to explain it?</p> <p>What means are being used?</p>	<p>CITIZEN ENGAGEMENT</p> <p>What form should citizen engagement take?</p>  <p>How are citizens involved?</p> <p>How do citizens learn about the output of the system?</p> <p>How do citizens challenge the output of the system?</p>	
<p>LEA ENGAGEMENT</p> <p>In which form should law enforcement agencies be included?</p>  <p>How are law enforcement personnel involved?</p> <p>Are there differences between sworn/non-sworn personnel?</p>			

Operational considerations

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.

TRANSPARENCY

Meaning

Transparency is a principle that involves making available clear, accurate and meaningful information about AI processes, decisions, technologies, capabilities and specific deployments 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.³⁰⁹ 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.

Materiality threshold

While the accessibility and intelligibility of timely information is a fundamental requirement for meaningful accountability, not all data in AI-led programmes will be relevant to accountability or of sufficient importance that its full and immediate publication will make a material contribution to accountability. The parameters for not publishing information and a mechanism whereby the information can be accessed if it becomes relevant (such as by way of legal challenge or complaint) should be identified **and published** in advance.

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.

Note on Human Right Impact Assessment guideline and Data Protection Impact Assessment

Transparency may be a requirement with regard to assessment procedures, but the prevailing orientation is for a limited level of transparency, usually restricted to the main findings of the evaluation results. In addition, there are cases in which full disclosure of the assessment results may be limited by the legitimate interests of the data controller, such as confidentiality of information, security, and competition. In this regard, the *Guidelines on Big Data* adopted by the Council of Europe in 2017 specify that the results of the assessment proposed in

the guidelines “should be made publicly available, without prejudice to secrecy safeguarded by law. In the presence of such secrecy, controllers provide any confidential information in a separate annex to the assessment report. This annex shall not be public but may be accessed by the supervisory authorities.”³¹⁰ Limits on transparency in relation to data subjects may be justified by the nature of the AI applications where information on their functioning needs to be protected to safeguard technical and operational methods (e.g., crime prevention, anti-fraud applications). However, in such cases, independent committees of experts may play a significant role in monitoring AI systems and Transparency should be ensured to enable their activity.³¹¹

Implementation guide

<p>RECIPIENTS</p> <p>Who needs to offer transparency? And to whom?</p>  <p>Define transparency for this purpose</p> <p>Who offers transparency and who receives or benefits from it?</p>	<p>DATASET SCOPE</p> <p>Consider the importance of the size, nature and source of the datasets being used and the criteria for algorithmic processes.</p>  <p>What protocols deal with errors and malfunction?</p>	<p>MEASUREMENT</p> <p>What processes and criteria are used to judge whether the principle of Transparency has been sufficient complied with?</p>  <p>How is freedom of information applied for AI?</p> <p>What legislation applies in the context of AI?</p>
<p>PUBLIC CONCERNS</p> <p>Are public concerns being addressed when making decisions about transparency?</p>  <p>What are the considerations for different sections of society?</p> <p>Have any local courts, tribunals, regulators raised concerns?</p> <p>How do citizens learn about or challenge the outputs?</p> <p>What are the power relationships between stakeholders?</p> <p>What impacts do they have on system risks and benefits?</p>	<p>MAXIMISATION</p> <p>Maximising transparency should be considered at all stages, from system development to results</p>  <p>Is the system monitored after the testing phase?</p> <p>Is the approach based on a presumption of disclosure?</p> <p>Is the monitoring at all times or in a timeframe / intervals?</p> <p>On which measures is the monitoring conducted?</p>	<p>RESTRICTIONS</p> <p>Are there any legal or sector-specific restrictions to achieving transparency?</p>  <p>What are the legal or sector specific restrictions?</p> <p>What other methods for transparency are available?</p> <p>Can decisions be taken via a different procedure?</p> <p>Are there additional confidentiality issues?</p> <p>How is confidentiality retained while ensuring transparency?</p> <p>How is confidentiality ensured with public trust/confidence?</p>
	<p>DELIVERY</p> <p>Ensure transparency is achieved in a timely, meaningful and appropriate way.</p>  <p>Document flow on data sharing (who and when)</p> <p>Are there data publication and sharing protocols?</p> <p>Who has responsibility for data sharing oversight?</p>	

TRANSPARENCY
Ensures availability and ready accessibility of information pertinent for assessing and enforcing accountability to all relevant stakeholders.

Operational considerations

Transparency is fundamental to achieving AI 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.

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.

Materiality threshold

As in any other sectoral partnerships it can be expected that an AI programme will have some existing relationships and connections between individuals, teams and departments. In particular there may be arrangements for IT support provided by the security practitioner where external SME and volunteer bodies are involved in scrutiny and consultation for the AI programme. Similarly, there are some highly specialised activities and expertise where the availability of individuals and organisations having the right qualifications, experience and security clearance may require the involvement of people with an existing or previous relationship with the practitioner. An assessment of the materiality of any existing connections, relationships and dependencies should be made and documented at the outset and its materiality to AI Accountability kept under review throughout the AI programme's life cycle.

Example of applicable laws

- National and European laws establishing statutory oversight roles and bodies.

Note on Human Right Impact Assessment:³¹² including guidance and professional practice published by colleges and professional bodies for the sector.

Note on Data Protection Impact Assessment: including guidance published by National Data Protection Authority.

Operational considerations

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.

Implementation guide

INDEPENDENCE

Guarantees that monitoring and enforcement are independent from the people and/or organisations that design, implement and/or use the AI system.

<p>OPERATIONAL DEFINITION</p> <p>Determine the nature and extent of independence in a practical sense.</p>  <p>Define the appropriate and applicable legal background</p> <p>Define the regulatory background</p>	<p>SCOPE</p> <p>If total Independence is not possible, which form and level of Independence is (in)appropriate?</p>  <p>Is the form of independence appropriate? How?</p> <p>Is the level of independence appropriate? How?</p>	<p>COMMUNICATION</p> <p>Have effective lines of interaction and communication with the oversight body been established?</p>  <p>Established lines of communication</p> <p>What are the internal communication protocols?</p> <p>Who develops, participates, oversees communication?</p>
<p>LIMITATIONS</p> <p>Determine potential practical or legal limitations to the overall aim of Independence.</p>  <p>Document the practical limitations</p> <p>Document the legal limitations</p> <p>Identify links between LEAs/AI partners and regulators</p>	<p>KNOWLEDGE ACQUISITION</p> <p>How will oversight bodies acquire the necessary specialist knowledge to be able to carry out informed, effective decision-making?</p>  <p>Detail processes to acquire specialist knowledge</p> <p>Resource availability for capacity building and staffing</p>	<p>PROCESS RELATIONSHIPS</p> <p>In case (non-AI specific) accountability processes are already in place, how are relationships between the different accountability processes regulated?</p>  <p>Describe integration with accountability processes?</p>
<p>REGULATORY RELATIONSHIPS</p> <p>How are relationships of the accountability oversight body regulated with pre-existing oversight bodies?</p>  <p>Map existing oversight bodies</p> <p>Define integration approach for oversight bodies</p> <p>Can regulatory bodies compel disclosure of data at any stage?</p> <p>Has the need for data disclosure been made clear?</p>	<p>COMPLETENESS</p> <p>Information being provided to the oversight body must be adequate for the purpose of accountability?</p>  <p>What information is needed?</p> <p>How is it documented and where is it stored?</p> <p>Who is responsible for this information?</p>	<p>COMPONENT PROCUREMENT</p> <p>If institution is procuring parts or elements of the system from third-parties how are they instituting appropriate governance controls?</p>  <p>List third-party vendors, contractors and developers</p> <p>Controls for accountability, traceability & auditability</p>
<p>LEA POSITIONING</p> <p>Does Independence exclude LEAs from accountability bodies?</p>  <p>Are LEAs included in the accountability bodies?</p>	<p>DATA PROCUREMENT</p> <p>If third-party data is being used in the production of the AI system, how are they instituting appropriate governance controls across the data lifecycle?</p>  <p>List third-party datasets ad owners</p> <p>Controls for accountability, traceability & auditability</p>	

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.

Materiality threshold

This Principle partly embodies the concept of materiality: if evidence is robust and relevant then it is likely to be of potential material value. The relevance and weight of any specific evidence to the overall programme will need to be considered and evaluated. Where robust evidence is regarded as being of only marginal or tangential relevance and weight, this will need to be identified and recorded to avoid the appearance of improper bias or selectivity in the same way as academic rigour is applied to research programmes.

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.

Note on Human Right Impact Assessment guideline and Data Protection Impact Assessment

Both HRIA and DPIA are evidence-based processes,³¹³ and all the three main steps involved (planning and scoping, risk analysis and assessment, mitigation and further implementation) must be documented and based on concrete evidence. In this sense, the main characteristics of the product/service, the legal context, the relevant rights holders and stakeholders, the rights potentially affected, the likelihood and severity of potential impacts, the measures taken and their effectiveness in addressing the risks must be properly analysed and documented.

Implementation guide

COMMITMENT TO ROBUST EVIDENCE

Ensures that mechanisms are in place that lead to robust evidence which forms the basis for the assessment and enforcement of AI systems and their usage.

<p>OPERATIONAL DEFINITION</p> <p>How to define and assess 'robustness', and who is responsible to determine 'robustness'?</p> 	<p>INTENTION</p> <p>For what purposes might the evidence be used?</p> 	<p>EVIDENTIAL CONSTRAINTS</p> <p>Is the evidence in its original form subject to legal or sector-specific constraints?</p> 
<p>What is the definition of robustness?</p>	<p>Document the potential use of the evidence</p>	<p>What are the legal or sector specific restrictions?</p>
<p>Document how robustness will be assessed</p>	<p>EVIDENCE CAPTURE</p>	<p>How is evidence managed for accountability & compliance?</p>
<p>Who has responsibility to determine robustness?</p>	<p>Are processes and procedures in place to allow the capture of the evidence in the required way?</p>	<p>Is it legally compliant and easily understood?</p>
<p></p>		<p>Is it compliant with other principles?</p>
<p>PROCESS MANAGEMENT</p>	<p>What is the legal background to processes and procedures?</p>	<p>Document how compliance is achieved</p>
<p>Are these processes and procedures documented and understood by those who need to know?</p>	<p>What is the process for documentation?</p>	<p>INTERPRETATION</p>
	<p>STORAGE AND ACCESS</p>	<p>How should non-AI experts learn to interpret AI outputs in the evidential context?</p>
<p>Map the flow of accountability</p>	<p>Is the evidence stored in a meaningful and accessible way?</p>	
<p>Map accountability responsibilities</p>		<p>Does the AI documentation provide explanations?</p>
<p></p>	<p>How is the data stored and how can it be accessed?</p>	<p>How is consistency of interpretation measured?</p>
<p>RESILIENCE</p>	<p>How is evidence protected from tampering?</p>	<p>Who checks for ongoing consistency?</p>
<p>Is the evidence sufficiently robust for these purposes?</p>		<p></p>
<p>Verify the evidential integrity</p>	<p></p>	<p></p>

Operational considerations

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.

ENFORCEABILITY AND REDRESS

Meaning

The principle of Enforceability and Redress requires that relevant oversight bodies and enforcement authorities have access to the necessary powers, means and mechanisms to respond appropriately to instances of non-compliance with applicable obligations by those deploying AI in a criminal justice context. A crucial aspect of this is to give effect to individuals' fundamental right to an effective remedy,³¹⁴ established at European Treaty level. However, there are also highly relevant 'internal' mechanisms for individual Enforceability and Redress such as professional standards in policing and criminal justice, codes of conduct, employment and other contractual arrangements. Enforceability and Redress in AI projects can also be achieved via standards that are set by regulators such as the national Data Protection Authority or Forensic Science Regulator. Non-compliance with these standards can result in substantial fines, reputational damage and exclusion from procurement exercises.

Materiality threshold

Internal and external measures for ensuring Enforceability and Redress are essential in giving the Principles 'practical effect'. However, not every shortcoming or departure from the agreed project variables has to be capable of enforcement and not every aspect of the project needs to be backed up with powers of compulsion. Where any peripheral shortcomings or shortfalls occur, it may be necessary to identify their significance in terms of **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.

Note on Human Right Impact Assessment: including guidance and professional practice published by colleges and professional bodies for the sector.

Note on Data Protection Impact Assessment:³¹⁵ including guidance published by National Data Protection Authority.

ENFORCEABILITY AND REDRESS

Ensures mechanisms are in place to enforce relevant obligations (legal, ethical, AP4AI) and recommendations of accountability oversight bodies as well as to guarantee implementation of remedies in case of negative impacts, consequences or grievances.

Implementation guide

<p>OBLIGATION</p> <p>Which obligations are capable of enforcement?</p>  <p>What are the legal obligations for enforcement?</p> <p>What are the non-legal obligations for enforcement?</p> <p>Who has discretion on the decision of enforcement?</p>	<p>COMPREHENSION</p> <p>Have steps been taken to ensure that the enforceability mechanisms are clearly understood?</p>  <p>Document that enforceability mechanisms are understood</p>	<p>HUMAN OVERSIGHT</p> <p>Is there continuous and effective human oversight over decisions made based on AI findings?</p>  <p>Who is monitoring the AI system</p> <p>Ensure decisions made are fully documented</p>
<p>FULFILMENT</p> <p>Who determines whether obligations have been fulfilled?</p>  <p>What and who is responsible for the obligations?</p> <p>What is the (internal) process and legal background?</p>	<p>CONFLICT RESOLUTION</p> <p>Has a conflict resolution and escalation process been identified?</p>  <p>Document the conflict resolution process</p> <p>Have all stakeholders signed up to the process?</p>	<p>ACCESSIBILITY</p> <p>Is information relating to obtaining an effective remedy clear, easily understood and accessible?</p>  <p>Verify understanding of effective remedy documentation</p> <p>How is the effectiveness measured and who is responsible?</p>
<p>SELECTION</p> <p>Which forms of redress will be chosen and how are they related to existing (national, international) redress possibilities?</p>  <p>Detail the forms of redress available</p> <p>What are the relevant sanctions and remedies?</p> <p>Document clearly what is the jurisdiction of each organisation</p>	<p>INTERVENTION</p> <p>When and for what reasons can regulators intervene?</p>  <p>Which regulators and stakeholders can intervene?</p> <p>Up to when can intervention happen before enforcement?</p> <p>Can intervention be an alternative for harm minimisation?</p>	<p>CONTESTABILITY</p> <p>Are there options for people affected by a decision to learn about the output of the automated system and to challenge predictions/ recommendations/ decisions influenced by the system?</p>  <p>How do citizens learn about / challenge the output of the system?</p>
<p>RESPONSIBILITY & INDEPENDENCE</p> <p>Who determines the appropriate level of redress?</p>  <p>Document the people and their roles</p> <p>Are those enforcing independent from those implementing redress?</p>	<p>HARM MANAGEMENT</p> <p>Are there internal responsibility procedures in place to address any unintended harm caused by the design, development or deployment of AI?</p>  <p>Document the processes to manage unintended harm</p>	<p>RECOURSE</p> <p>Are there any complaints or appeal procedures, or any type of recourse available for any harm caused by AI in the process of decision making?</p>  <p>Provide a process for recourse from AI decisions</p> <p>Document how complaints and appeals are managed</p>
<p>EXTERNAL OVERSIGHT</p> <p>Are there any internal or external monitoring, auditing or oversight procedures for evaluating the use of AI and assessing their impact on users or other individuals / groups?</p>  <p>Document external monitoring, auditing and oversight</p> <p>Document how external monitoring processes work</p>	<p>EXTERNAL OVERSIGHT</p> <p>Are there any internal or external monitoring, auditing or oversight procedures for evaluating the use of AI and assessing their impact on users or other individuals / groups?</p>  <p>Document external monitoring, auditing and oversight</p> <p>Document how external monitoring processes work</p>	<p>REVERSIBILITY</p> <p>Is the harm of a wrong decision by AI system fully reversible?</p> 

Operational considerations

Compliance with existing legal obligations 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.

COMPELLABILITY

Meaning

Compellability 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 AI accountability process by regulating the timely provision of relevant, up to date and accurate information in an intelligible format. Linked closely with Enforceability and Redress, this Principle will be greatly enhanced if it is supported within the terms of any contracts and Data Sharing Agreements.

Materiality threshold

As with Enforceability and Redress, Compellability does not have to be available for each and every facet of the project. There may be minor deviations from, for example, timescales for the provision of AI for the project or slippage in terms of dates, budget reporting and so on. It would be unrealistic and unconstructive to insist on Compellability mechanisms in every such instance and an assessment may need to be made as and when tangential or minor matters arise during the project's lifecycle.

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.

Note on Human Right Impact Assessment: including guidance and professional practice published by colleges and professional bodies for the sector.

Note on Protection Impact Assessment: including guidance published by National Data Protection Authority.

Implementation guide

<p>OVERSIGHT CAPACITY</p> <p>The oversight body's role and authority, functions and powers should be determined</p>  <p>Define the body's role, authority, function & powers</p> <p>Document the degree of information access required</p>	<p>REQUIREMENTS</p> <p>What process is in place to clarify and explain what is required, in respect of information and access?</p>  <p>Document the processes for providing information</p> <p>Document the process for allowing access</p>	<p>NOTIFICATION</p> <p>What mechanisms are used to inform LEAs of and conduct actions related to compellability?</p>  <p>Document how LEAs are informed</p> <p>What processes apply in relation to compellability?</p>
<p>OVERSIGHT POWERS</p> <p>On what grounds can oversight bodies interrupt, interrogate or compel LEAs or programme partners either directly or via national bodies such as regulators?</p>  <p>Document the grounds for invention</p>	<p>INFORMATION SECURITY</p> <p>Have legal and sector-specific obligations in respect of information security been complied with?</p>  <p>What are the information security obligations</p> <p>What are the legal or sector specific restrictions?</p>	<p>NON COMPLIANCE</p> <p>Have the sanctions or consequences of non-compliance been clearly communicated?</p>  <p>Document the consequences of non-compliance</p> <p>How are the sanctions communicated and to who?</p> <p>What conditions apply and how were they determined?</p>
	<p>SECURITY AND SAFEGUARDS</p> <p>What security measures and other safeguards are in place in respect of the provision of information?</p>  <p>Document security on the provision of information</p>	

COMPELLABILITY

Requires that obligations are in place that ensure LEAs provide oversight bodies with access to required information, AI systems or individuals.

Operational considerations

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.

EXPLAINABILITY

Meaning

Explainability requires those using AI to ensure that information about this use is provided in a meaningful way that is accessible and easily understood by the relevant participants and audiences. This Principle 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. Given some of the well-publicised concerns around the extent to which AI algorithms are understood or even capable of explanation, this Principle is of significance in terms of public accountability. It might be that a basis level of Explainability is expressly built into contractual agreements with designers and providers.

Materiality threshold

There may be technical elements of an AI programme which, while relevant to the effective operation and functioning of the technology itself, are not sufficiently material **to the accountability considerations as set out in the other Principles**. The potential for confusion, mistrust and even suspicion that is inherent in security-related AI programmes makes the assessment of materiality a critical element when it comes to Explainability. The default position should be that every element covered by the Principles ought to be capable of explanation to the interested citizen and that any areas of 'inexplicability' should be very much the exception.

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, ability to provide a defence in criminal proceedings and the prevention of arbitrary decision-making.

Note on Human Right Impact Assessment guideline and Data Protection Impact Assessment

Although both HRIA and DPIA are not Explainability tools and are usually subject to limitations regarding the transparency of the findings, they can contribute to the internal process of AI design for a better understanding of the potential impacts of AI in relation to Human Rights and fundamental freedoms. In addition, access by supervisory authorities to the results of these assessment tools and,

where applicable, access granted to auditing bodies or rightsholders further contribute to better explaining the functioning of AI from the point of view of its impact.

Implementation guide

<p>SCOPE OF APPLICATION</p> <p>For which aspect(s) of AI or AI usage is Explainability relevant?</p>  <p>Document AI areas relevant to explainability</p> <p>Document justifications for areas that are excluded</p> <p>Provide transparency of processes and outcomes</p>	<p>COMMUNICATION STRATEGIES</p> <p>Are clear communication strategies in place that account for different needs of individuals and groups?</p>  <p>Is the nature and type of information considered?</p> <p>Which processes ensure effective communication?</p> <p>Is there dedicated resources for communication?</p> <p>Do these strategies align with the intention of the AP4AI?</p> <p>Has a publication strategy been developed?</p>	<p>EFFECTIVENESS</p> <p>How is the effectiveness of this principle measured? ?</p>  <p>How is effectiveness measured?</p> <p>What factors have been taken into account?</p> <p>What dependencies impact explainability availability?</p>	<p>EXPLAINABILITY</p> <p>Ensures that AI practices, systems and decisions can be explained.</p>
<p>FULFILMENT</p> <p>How to determine whether Explainability has been satisfied? Who judges whether Explainability has been satisfied?</p>  <p>Document how explainability is satisfied</p> <p>What are the assessment bodies and mechanisms?</p> <p>Is meaningful information about decision logic provided</p> <p>Is meaningful information provided on consequences?</p> <p>Can staff interpret and challenge information?</p> <p>Can stakeholder groups understand the information</p>	<p>RISKS AND CONSEQUENCES</p> <p>Is there clear understanding of the significant risks and consequences of not complying with this principle</p>  <p>Are relevant risks and consequences documented?</p> <p>How have mitigation measures been documented?</p>	<p>REVIEW MECHANISM</p> <p>Have mechanisms to facilitate reviews, challenges and complaints been established?</p>  <p>Document review, challenge and complaint mechanisms</p> <p>Communicate the procedure for contesting decisions</p>	

Operational considerations

The diversity of relevant stakeholders in the AI 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.

CONSTRUCTIVENESS

Meaning

Constructiveness 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.

Materiality threshold

Accountability means that there will be occasions where reports and findings are appropriate, even though they are highly critical of, and potentially challenging to policing and the justice system. The Principle of Constructiveness must not be allowed to dilute the proper accountability mechanisms or to have an adverse impact on the other Principles (for example Transparency, Commitment to Robust Evidence, Enforceability and Redress) – at the same time it aims to reduce the misuse of data and research reports in furtherance of extreme or malign activity against relevant organisations.

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.

Note on Human Right Impact Assessment guideline and Data Protection Impact Assessment: see the considerations made on stakeholder involvement and participation with regard to Pluralism.

Implementation guide

<p>MECHANISMS AND SAFEGUARDS</p> <p>What are mechanisms to safeguard Constructiveness in discussions and negotiations?</p>  <p>Make written guidance for discussion available</p> <p>Provide staff with adequate knowledge and training</p> <p>How is robust enforcement balanced with improvement?</p>	<p>STAKEHOLDERS</p> <p>Who are the stakeholders that need to be involved?</p>  <p>List the stakeholders involved</p> <p>Has a stakeholder briefing on this principle been conducted?</p> <p>Verify their understanding of how it relates to accountability</p>	<p>DISCONNECT</p> <p>How to handle actors that fail to adhere to a basic foundation of Constructiveness?</p>  <p>Describe the process for handling unwilling actors</p>
<p>RISK MANAGEMENT</p> <p>Have specific risk(s) in relation to this principle been added to the risk register</p>  <p>Ensure risk register captures all specific risks</p> <p>Document the process for monitoring of risk</p>	<p>COMMUNICATION</p> <p>How are communications managed internally and externally to ensure this principle does not dilute accountability and transparency?</p>  <p>How can stakeholders challenge the application of the principle?</p>	<p>ENGAGEMENT</p> <p>Has an independent spokesperson been identified who can address key accountability issues arising at each stage of the project?</p>  <p>Nominate the spokesperson and clarify their responsibilities</p>

CONSTRUCTIVENESS

Ensures a dialogical process between law enforcement agencies and judicial actors, and those performing accountability functions, that is enabling and responsive.

Operational considerations

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.

CONDUCT

Meaning

Conduct requires that AI practices should always be able to stand up to scrutiny by the public and other bodies, by adhering to sector-specific principles, professional standards and expected behaviours relating to conduct within a role, which incorporate integrity and ethical considerations. The Conduct expectations and relevant standards for individuals and organisations involved in activities relating to AI must be expressly identified in advance along with the relevant means that will be used to hold them to account.

Materiality threshold

The processes for identifying and addressing areas of conduct internally and externally will involve an assessment of materiality and proportionality. To that extent this Principal incorporates a materiality threshold, whether that is for alleged breaches of criminal and civil law, professional codes of conduct and domestic frameworks set by industry regulators such as national data protection authorities. Nevertheless, there may be situations where consideration of the relative contribution to a 'wrong' of an organisation or individual associated with the AI project needs to be considered.

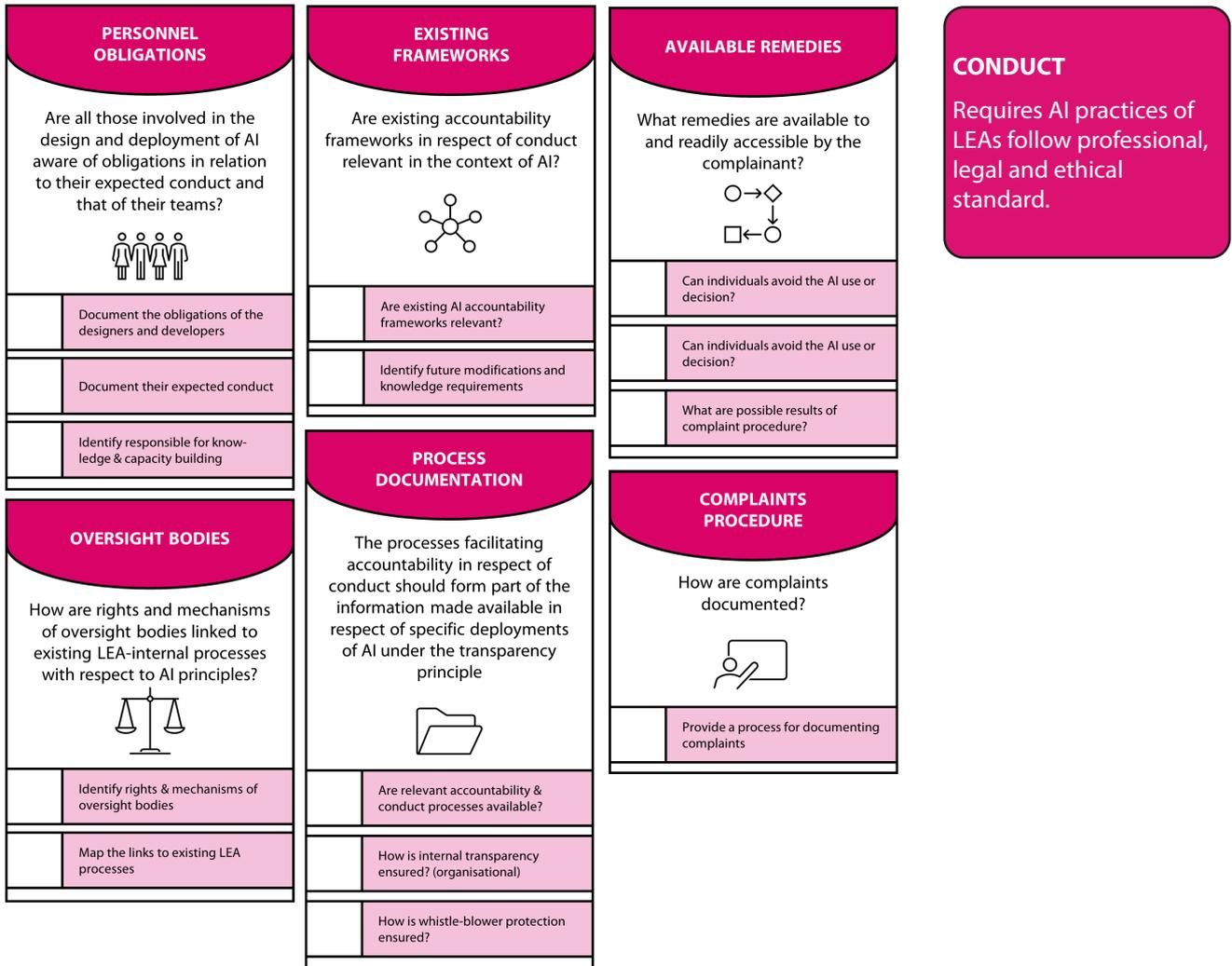
Examples of applicable laws

- 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.
- Professional standards.

Note on Human Right Impact Assessment guideline:³¹⁶ including guidance and professional practice published by colleges and professional bodies for the sector.

Note on Data Protection Impact Assessment: including guidance published by National Data Protection Authority.

Implementation guide



Operational considerations

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. Where partners in the AI ecosystem are from jurisdictions with different forms of state rule and/or have different values, there may be a requirement for closer scrutiny and review mechanisms and even barriers to entry into AI programmes involving accountable policing organisations. Consequence for non-adherence may vary according to sector, ranging from internal disciplinary proceedings to formal professional sanctions and even proceedings before courts or tribunals.

LEARNING ORGANISATION

Meaning

The Principle Learning Organisation 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.

Materiality threshold

There will be many learning opportunities arising from AI programmes, not all of which will be relevant or sufficiently conclusive. It is important to identify the key areas at both a prospective and summative stage (*ex ante* and *ex post*) using the robust evidence generated by the programme and the preceding Principles. Conducting a post-project evaluation of 'what worked and why' will assist in identifying the material contributions to learning for the organisation, some of which will probably include avenues for further research, evaluation and ongoing review.

Examples of applicable laws

- 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.

Note on Human Right Impact Assessment and Data Protection Impact Assessment

Both in the HRIA and DPIA require specific skills and expertise in those who carry out them. Moreover, the first stage of these processes (planning and scoping) requires a contextual analysis of the relevant issues, learning the key elements of the concrete framework in which AI is used and its dynamics.

Implementation guide

<p>UNDERSTANDING</p> <p>How do security practitioners learn about/are informed about aspects that need to be adapted?</p>  <p>How do security practitioners learn?</p> <p>What resources are available?</p>	<p>KNOWLEDGE MANAGEMENT</p> <p>How is learning codified to ensure it remains available, replicable and can spread within the organisation/ sector?</p>  <p>What is the process for documenting learning?</p> <p>How is learning disseminated through the organisation?</p> <p>Who are the recipients of the learning?</p>	<p>RESOURCE AVAILABILITY</p> <p>Are sufficient resources in place to enable and sustain the learning?</p>  <p>Document the available resources for learning</p> <p>Who is responsible for resources and materials?</p>
<p>STAKEHOLDERS</p> <p>Is learning only needed for security practitioners or are other groups equally required to make adjustments?</p>  <p>Document all the groups that are involved in learning</p>		<p>EVALUATION</p> <p>Are sufficient resources in place to enable and sustain the learning?</p>  <p>Document the available resources for evaluation</p> <p>Who is the person responsible for evaluation?</p>

LEARNING ORGANISATION

Ensures the willingness and ability of organisations and people to change current AI practices based on new knowledge and insights.

Operational considerations

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.

APPLICATION SCENARIOS – USE CASE EXAMPLES

The use cases in this section are selected to cover a broad range of security challenges given the increasing use of AI by internal security practitioners and, most importantly, citizen demands for the implementation of AI in these contexts (see [section on Citizen consultation](#)). The motivations for their choice are underpinned by the results from the citizen consultation. Following the model for the AI Accountability Agreement (AAA), we put forward three example problem scenarios (**the context**), potential applications of AI in these settings (**the scope**) at different points of the investigation lifecycle, the broader considerations for AI deployment (**the methodology**) and the specific requirements for the management of the AI-related applications (**accountability governance**). The below discussions are not intended to be comprehensive, but to give a flavour of different considerations at each stage of developing an AAA. This approach demonstrates and ensures that the Framework and the AAA is fit for the complexities in real-life applications of AI in the internal security domain.

In the next cycle of the AP4AI Project, we will apply and re-validate the AAA and the Accountability Principles Implementation Guide through further validation and contextualisation efforts (e.g., co-creation workshops) governed by these use cases. The actual scenarios will be developed further with JHA³¹⁷ partners and wider stakeholders so that the AP4AI Implementation Guide and its supporting software tool can be used with kitemaking quality of AP4AI. All 12 Principles will be applied through the four elements of the AAA.

UC1: Counter-terrorism – online terrorist generated content

Motivation

Terrorism has a major impact on all aspects of society. Terrorist attacks have unfortunately been an enduring presence across Europe as can be seen from the 2005 London Bombings to the 2015 Paris attacks. We do not see attacks of this magnitude occur often. Yet, according to a recent Europol report³¹⁸, in 2019³¹⁹ alone a total of 119 foiled, failed and completed terrorist attacks were reported by 13 EU Member States, with 1004 individuals being arrested on suspicion of terrorism-related offences in 19 EU Member States. As a result of these attacks, ten people died in the EU and 27 people were injured. An analysis of recent terrorist activities reveals a growing use of ICT, while the arrangements for identifying and monitoring those known to be involved in terrorist activity utilise biometric surveillance capabilities. Consequently, investigation and prevention is increasingly dependent on fast and reliable analysis of large quantities of data. AI tools and machine learning approaches have the potential to significantly enhance the capacities of LEAs to carry out such data-focused investigations (e.g., supporting data and intelligence collection, processing, analysis and cross-referencing). Therefore, AI can significantly improve the efficiency of counter-terrorism agencies in Europe, and ultimately minimise the threat of terrorism in the EU.

Context

One of the unique challenges counter-terrorism units face is the need for proactive monitoring of online platforms for terrorist content to allow them to intervene prior to an attack. Terrorist groups are becoming increasingly decentralised and generally do not use a single platform to spread propaganda and communicate. Instead, terrorist actors utilise all forms of social media, surface and dark web forums allowing them to reach members of terrorist groups, sympathisers and those at risk of radicalisation. Only by monitoring and analysing the full range of these platforms, can analysts gain insights into the material to prevent attacks and identify those individuals and groups involved. Technological advancements mean that there is now potential for AI tools to support across multiple facets of terrorism investigations; particularly allowing them to monitor and identify pertinent data from the vast troves of information posted online following a targeted and intelligence-led approach.

Scope

AI applications can be set to automatically monitor and extract information from multiple online sources whilst also identifying previously unknown sources of content. Starting from known and investigator validated online sources, the ability to rapidly extract and identify updated content releases an investigator from arduous and manual checking and downloading of such content.

Information extraction, text and multimedia analysis applications can be put to work to obtain key pieces of information (names, dates, locations, financing, threats, etc.) based on generalised information models and specialised domain expert created taxonomies that allows for quick removal of irrelevant content and for relevant content to surface quickly. AI applications can also recognise individuals, usernames and avatars placing all information relevant to a specific person of interest in one place. More sophisticated text analysis approaches can detect escalating intent or radicalisation or engage in discussion with potential suspects or at-risk individuals through the use of bots.

Finally, the construction of networks of relationships between and across platforms resolving profiles to specific individuals and groups and making predictions based on acquired information allows for important groups of chains of individuals to emerge and are not blinkered by any preconceived ideas from the investigator. Furthermore, all these AI-supported activities can be performed at a higher rate than at present, provide automated alerting and prevent pertinent information being lost in a sea of data by surfacing the most relevant and actionable information in a timely manner.

Data at this stage may take the form of specific source URLs and accounts, extracted web pages, social media posts, online profiles, salient information such as names, organisations, locations, and the relationships between them. Example stakeholders in the setup of such an operation will include law enforcement personnel, platform owners whose services host the content, AI-tool developers and legitimate users of the platforms, in addition to the suspects themselves.

Methodology

The above types of applications of AI may need to be developed through in-house software development, customisation of an existing off-the-shelf product or through procurement of an external system. Taking the example of in-house software development, an appropriate software development process should be defined (e.g., an agile process) that is in line with the organisation's other existing IT processes. The deployment of AI introduces additional considerations around the management of training and testing data, the labelling of data (e.g., types of terrorist groups, types of content or activities) and the processes for the update and management of the AI models to remain current with evolving terrorist activities, and specific customisations in relation to the information sources data should be collected from. Furthermore, integration with other intelligence databases (both internal and external to the organisation) and how such data is imported and exported to those systems as part of the AI development and output needs to be defined. Similarly, export to existing analytical systems and the process for the refinement of the models based on analytical outputs must be clear from the outset.

Governance

The use of AI-enabled tools to obtain and analyse information from online platforms may raise challenges in terms of incidental processing of non-suspects' personal data, and the automation of monitoring tools may infringe on the platform's terms of use and privacy policies. Traceability from data source to outcome, oversight mechanisms for source selection and prioritisation approaches must be considered here. For advanced analytics, such as network analysis and prediction the underlying process for reaching recommendations should also be explainable. For regular online monitoring processes to verify the continual applicability of the source URLs and the relevance of the extracted data should be defined in advance of the data capture.

UC2: Child sexual exploitation – Obtaining and sharing CSEM

Motivation

The detection of perpetrators involved in child sexual exploitation (CSE) is becoming an increasing challenge for LEAs. The CSE material that is shared by these individuals is often masked using virtual private networks (VPN) and other encryption tools, which causes a problem for LEAs in identifying the individuals sharing the material as well as their locations. Even for those who share the material freely online without hiding their identity, the sheer amount of reports of child sexual exploitation material (CSEM) makes it difficult for LEAs to handle the information and deal with it in a timely manner. Alongside this, LEAs have to handle a large number of reports of sexual grooming and communication of children online, with a recent Europol report suggesting there has been a huge increase of reports since the start of the COVID-19 pandemic.³²⁰ One consequence of this increase in workload is the impact on human investigators and their emotional wellbeing after continued exposure to graphic and disturbing material.³²¹

Context

Exploitative individuals will actively seek out conversations with children, with the aim to have the child share sexually explicit material of themselves either through coercion or grooming. This can then lead to content being shared online or within peer-to-peer networks without the child's knowledge or threats being made to the child that the material will be shared with their peers unless further content is produced. AI applications can be deployed at numerous stages of this process by platforms and services facilitating the communication, by NGOs and other independent referral agencies as well as by LEAs. Given the potential harm to victims, early intervention is key, and AI can be deployed to do this unobtrusively and without infringing on the rights of the individual rights of platform users before an incident as well as in the investigation phase.

Scope

Platform providers may deploy AI applications in the forms of age detection software to identify communication between younger and older individuals or use forms of real-time conversational monitoring to alert users about inappropriate and high-risk communication activities. Platform providers may also be able to intercept CSE material uploaded to their platforms to prevent it being shared further, using multimedia classification or categorisation methods to detect age profiles, nudity or sexual acts. Furthermore, comparison against existing CSEM hash databases can also prevent existing material being reshared. AI can support this feedback loop by also empowering LEAs to add CSEM hashes to existing databases in a rapid manner and through automatic processing based on smaller amounts of detected content (e.g., high confidence for automatically classified images to contain CSEM from a sequestered hard drive or cloud storage with thousands of images after a small amount have been manually verified).

Post-incident, AI applications can support investigations in the analysis of large volumes of conversations between the perpetrator and the victims and the detection of suspicious content related to coercion. Similarly, NLP can be deployed to compare writing styles across multiple authors to identify where the same perpetrator may be posting under different profiles. It can also reduce the burden on investigators in terms of having to read or watch significant volumes of distressing material as well as utilising author or speaker recognition to match perpetrators across investigations.

Methodology

In the domain of CSE, many tools already exist to support the analysis and investigation of content. It is therefore important that applications of AI leverage this existing data and avoid reinventing the wheel. Use of technologies such as PhotoDNA³²² and other available hashing approaches and databases should be considered within the development process. Similarly, data providers (from personal reports to social platforms and organisations that operate as clearing houses, e.g., NCMEC³²³) all likely provide data in different formats with different standards of CSE classification and different terminology. Thus, prior to development, resolving and defining a standardised approach to classification of CSE material is essential to ensure the validity of both the models and the

outputs. A key component of defining the protocols, models, semantics and assumptions will be accurately describing and documenting the boundaries of each classification and the reasoning behind the approach.

Governance

Dealing with CSE related crimes can be distressing for all stakeholders involved in the case. The deployment of AI to support the detection and investigation of CSE-related activity should be carefully monitored for coverage (so platforms and LEAs are aware of what is not being detected) and particular (types) of CSEM are not disproportionately disadvantaged in their investigation and detection. A governance structure should be created in support of training, documentation and particular AI development strategy (e.g., Federated Learning or sandboxing when the no LEAs are involved) to address accountability. Furthermore, as with many applications of AI, NLP still has limitations in terms of what forms of languages (e.g., nuance, sarcasm, joking, slang) can be detected as well as differences in performance across languages, these should also be factored in and documented with the models to ensure consistent interpretation of the outputs.

UC3: Serious and organised crime – weak signal and crime prediction modelling

Motivation

In complex investigations of serious and organised crime, investigators must process extensive amounts of data relating to each aspect of the offence. Although there are some technical measures available to aid this, most of the work is done manually to ensure accuracy. These investigations often involve multiple suspects and other data points such as phone numbers, locations, transactions, and goods. Therefore, the ability to analyse data quickly is imperative to bringing the offenders to justice before further crimes are committed.

Context

Due to the vast amount of data and information available to LEAs, AI supported analysis can provide valuable support in identifying new and emerging trends, with the aim of alerting the authorities of 'weak signals' such as any new offence that is identifiable by a specific modus operandi and whose frequency has suddenly increased. An example of a 'weak signal' analysis could be the changing landscape of drugs trafficking. If a country were to see a rise in cases related to the use of a particular drug, weak signals can process the use of that word over a given period. The aim is to improve the responsiveness of LEAs in the face of ever-changing crime patterns and to better protect citizens to prevent them from falling victim to these new threats. Furthermore, an AI system may also propose countermeasures to address the threat, based on historical incidents of the same nature. To this extent, AI models can be trained on historical data of similar cases to assist LEAs in establishing the best approach to address an issue.

Scope

The application of AI to detect and monitor emerging serious and organised crime threats can be applied to various parts of the analysis process. In the context above, there is a greater emphasis on utilising existing internal investigation information to detect emerging patterns within that data. Therefore, there must be a process to automatically ingest and analyse newly added investigation data to scan it for relevant emerging trends. AI can be applied to problems such as keyword analysis as a timeseries to monitor against existing baselines for new and emerging terminology. Applications of AI can also be used to infer links between people, crimes, organisations and other pertinent data fields. Techniques such as clustering, graph analysis and pattern matching can flag when similar combinations of variables are showing up based on historical data and provide predictions using spatio-temporal crime models or uncover hidden patterns in underlying data that bring new insights into combatting organised crime.

Methodology

Developing analysis and predictive models is particularly precarious as the aim is to constantly uncover new information. Therefore, when defining the methodology for developing such applications of AI is important to accurately define the available data fields and create a taxonomy or knowledge base that ensure key terms and known indicators are documented and aligned. This is especially important for organised crime groups as they often are involved in multiple types of crime and being able to track information that could be siloed in cases about human trafficking when they are also involved in drug and gun crime is essential to ensure that all similarities across cases are considered. Given this poly-criminality, the other important facet of AI model development is laying down the approaches for model retraining as data and *modus operandi* evolve. For example, what are the defined intervals for model updates and retraining and how are the 'old' models stored and documented given specific intelligence may rely on the output of a specific model at a specific point in time. Storage of old models may also give rise to data retention considerations, as well as how and to when they should be deployed to the analytical environment.

Governance

Crime prediction modelling and weak signal Interpretation are powerful tools. However, the weaker the signal and the underlying dataset, the greater the risk that individual actions that go beyond general trend monitoring, and impact non-criminals or that incidental or spurious links are identified. Especially, algorithms and predictions developed and trained on existing investigation data could be unrepresentative of the wider prevalence of certain organised crime if it is underreported. Similarly, new trends may only show up investigatory data at certain thresholds therefore augmenting this data with information from other sources or utilising investigator knowledge is also key. Furthermore, prediction of potential criminal threats is notoriously difficult, therefore clearly thresholds on confidence of predictions, explainability and traceability mechanisms, transparency and oversight should all be clearly defined prior to any analytical activities.

The above three case studies will provide the baseline and framing protocols for AP4AI's forthcoming efforts for refinement, validation and contextualisation. They will steer the discussions with experts and support the validation and evaluation stages of the AP4AI Project. It is envisaged that further activities involve co-creation workshops with a diverse set of stakeholders from JHA and domain experts from law, AI, policing, judiciary, human rights, ethics and industry following on from the successful expert consultations in Cycle 1. The application of the 12 AP4AI Principles will be tested and validated using realistic problem statements and challenges. The outcome of these efforts will contribute to the next iterations of the AP4AI Framework and its associated components.

NEXT STEPS

This report represents a substantial set of outcomes from the AP4AI Project. It provided a critical commentary on existing bodies of knowledge and the wide spectrum of legal, legislative and policy documents, reported on initial result of the citizen consultation and finally outlines a blueprint for the implementation of AP4AI. The latter serves as a foundation for the upcoming activities towards the realisation of AP4AI's vision.

In the upcoming AP4AI activities, the project will provide:

1. Further validation and instantiation of the AI Accountability Agreement using real examples and challenges of internal security practitioners
2. In-depth analysis of the citizen consultation as an evidence-based instrument and provision of policy briefing to internal security stakeholders including EC agencies
3. Extension of use cases and application scenarios (AI deployment) into:
4. Investigation of CSE and categorisation of CSEM (Child Sexual Exploitation Materials)
 - Investigation of cyber-dependent crime
 - Identification and prediction of serious and organised crime activities including cross-border issues
 - Detection of harmful internet content such as terrorist generated internet content
 - Protection of public spaces and communities
 - Investigation of terrorism (including CVE) related offences
 - Investigation and prosecution of financial crime (e.g., money laundry)
 - Procurement of AI solutions by internal security practitioners
 - Research and development for AI either by the LEA or a 3rd party intended to create the solution to be deployed for the internal security domain
 - Identification of a future set of use cases
5. Further validation and contextualisation of AP4AI through combined methods such as expert inputs, focus groups and co-creation workshops
6. Development of a software application as a supporting mechanism for the implementation of AP4AI
7. Trainings and policy briefings for the internal security community
8. Extensive dissemination of project results and engagement with EU-funded projects, including ongoing and future research projects on AI

APPENDIX A: DETAILS ON DEVELOPMENT OF AP4AI PRINCIPLES

(Cycle 1 methodology)³²⁴

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

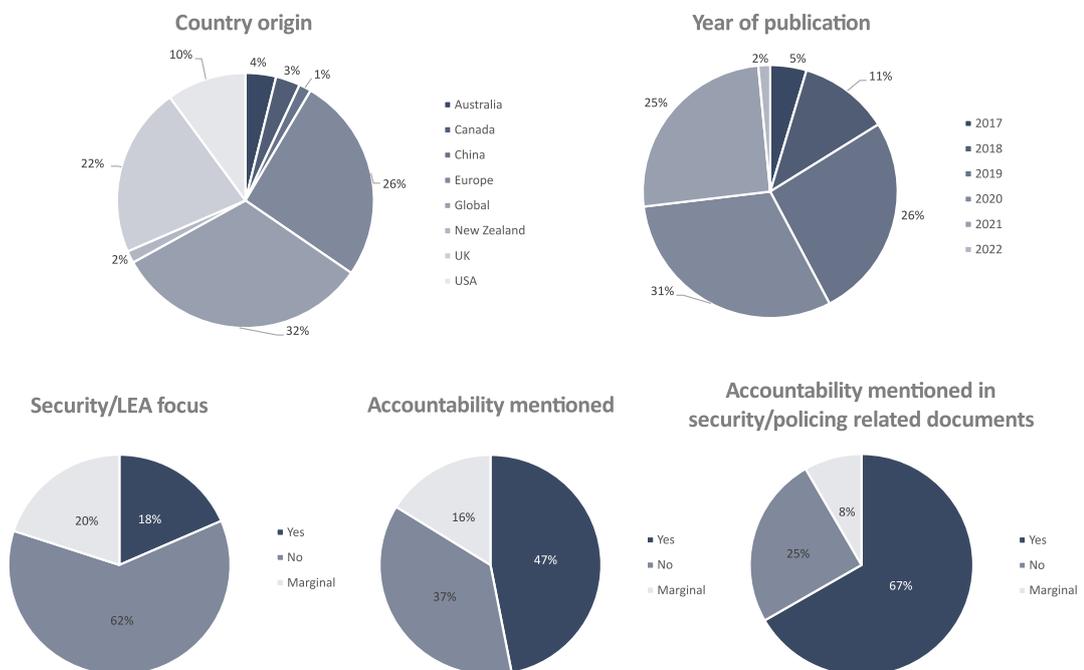
To ensure that AP4AI work and results are cognisant of as well as able to relate to and reflect latest developments, a comprehensive review of existing documents and reports was conducted. The selection of documents was purposefully broad to guarantee an expansive search. The following criteria were 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*: document 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 A1 in [section Collection of pre-consultation input](#) for an overview of the principles).

Figure A1 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.³²⁵ 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.³²⁶

Figure A1: Summary of relevant meta-information of the reviewed documents

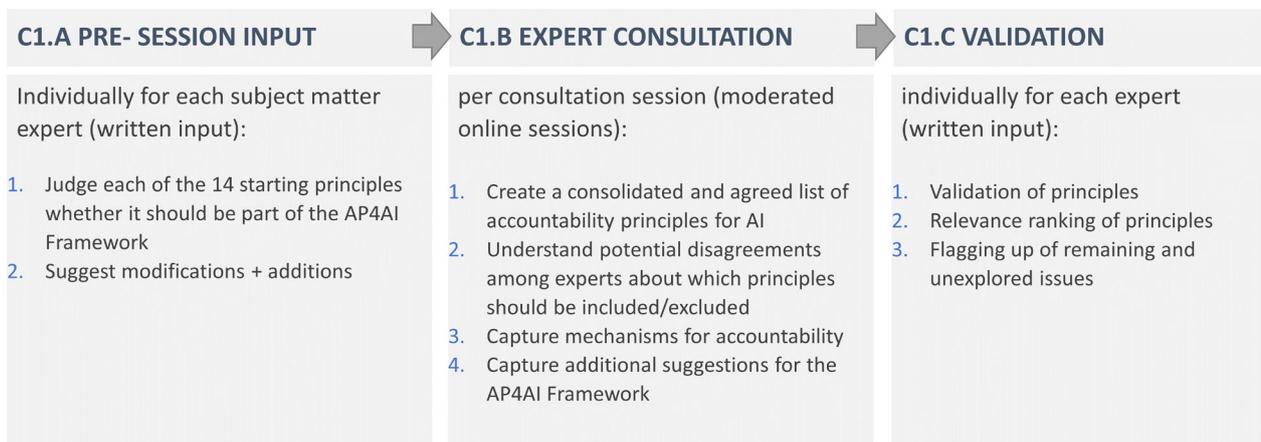


Subject matter expert consultations

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 validation forms.

Figure A2: 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)³²⁷ plus the principle of Trustworthy AI put forward by the European Commission's High-Level Expert Group on AI.³²⁸ 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 A1 provides an overview of the 14 starting principles as well as simplified definitions.

Table A1: 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 complaints³²⁹ 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.³³⁰
- 11. Recognition and Reason:** aims to facilitate 'participatory space' and encourage authentic public scrutiny.³³¹
- 12. Commitment to Robust Evidence and Independent Evaluation:** recognises that deliberations need to be informed by robust evidence and rigorous, independent evaluation

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 A3 for an illustration). They were further asked to add AI Accountability Principles they thought were missing.

Figure A3: Excerpt of the pre-consultation template

Principle (listed in random order)	A: Include as is	B: Include but needs adaption	C: Do not include	Explanations If B: explain adaptation If C: explain why
<p>Universality</p> <p>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</p>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

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 AP4AI 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.³³² Session length was capped at 2 hours. All expert consultation sessions were recorded and transcribed verbatim to allow detailed content analysis.

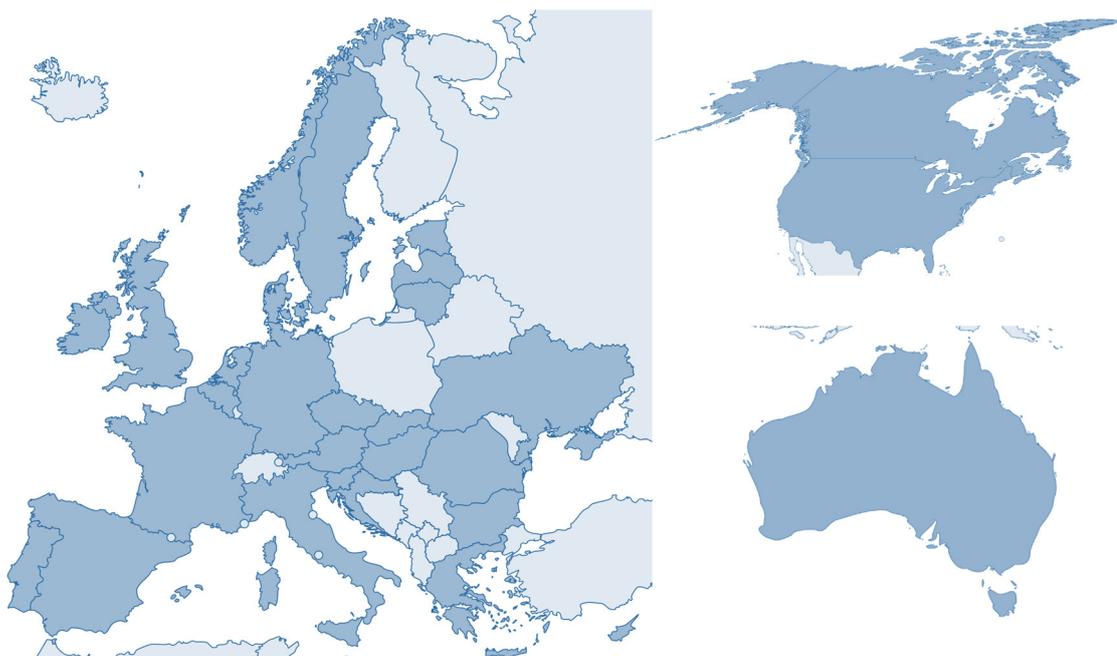
Expert inputs collected

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)
- **14/07/2021: 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 A4 indicates the countries with participation in Cycle 1.

Figure A4: 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 A2. 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 A2 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 A2). 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 A2: 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. 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.

ENDNOTES

- a In defining 'internal security' we follow the Internal Security Strategy 2010-2014 which was endorsed by the European Council: "The concept of internal security must be understood as a wide and comprehensive concept which straddles multiple sectors in order to address these major threats and others which have a direct impact on the lives, safety and well-being of citizens, including natural and man-made disasters such as forest fires, earthquakes, floods and storms." Internal Security Strategy 2010-2014. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2014\)542180](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2014)542180)
- b See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 1 Europol. (2021). Internet Organised Crime Threat Assessment 2021. https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2021.pdf
- 2 Trend Micro Research. (2020). Malicious Uses and Abuses of Artificial Intelligence. https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf
- 3 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.
- 4 For example, see proposed EU AI Act, Art 3,17 and 38; <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- 5 Interpol and UNICRI. (2020). Towards Responsible Artificial Intelligence Innovation. <http://www.unicri.it/towards-responsible-artificial-intelligence-innovation>
- 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. <https://www.legislation.gov.uk/ukpga/2011/13/contents/enacted>
- 8 See Akhgar et al. (2022). AP4AI Summary Report on Expert Consultation. <https://www.ap4ai.eu/node/6>
- 9 The AP4AI Principles are defined in the AP4AI Summary Report on Expert Consultations (<https://www.ap4ai.eu/node/6>). This report also outlines the overall project methodology and the activities leading to the creation of the AP4AI Principles.
- 10 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>
- 11 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 12 Schedler, A. (1999). Conceptualizing Accountability, in: A. Schedler et al. (eds), *The Self-restraining State: Power and Accountability in New Democracies* (pp. 13-28).
- 13 Ibid.
- 14 Thomas Reuters Practical Law. (2021) Accountability Principles. <https://uk.practicallaw.thomsonreuters.com/w-014-8164>
- 15 E.g., Duff, R. A. (2017). Moral and Criminal Responsibility: Answering and Refusing to Answer. <https://ssrn.com/abstract=3087771>
- 16 <https://www.hrw.org/world-report/2022/autocrats-on-defensive-can-democrats-rise-to-occasion>
- 17 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.
- 18 The overall project approach, together with a description of Cycle 1 activities can be found in [Appendix A](#). The text is replicated from Akhgar et al., (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6> for easier reference.
- 19 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 20 The full details of the activities in Cycle 1 are described in AP4AI Summary Report on Expert Consultation (<https://www.ap4ai.eu/node/6>) and can also be found in [Appendix A](#).
- 21 E.g., Deloitte Insights. (2020) Government Trends 2020. What are the most transformative trends in government today? Deloitte Center for Government Insights. https://www2.deloitte.com/content/dam/insights/us/articles/government-trends-2020/DI_Government-Trends-2020.pdf
- 22 Vaio, A.D., Hassan, R., & Alavoine, C., (2021). Data intelligence and analytics: A bibliometric analysis of

- human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change*. 174, 1-17.
- 23 European Commission. (2018). 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. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>
 - 24 House of Lords. (2020). AI in the UK, no place for complacency. <https://committees.parliament.uk/committee/187/liaison-committee-lords/news/138009/no-room-for-government-complacency-on-artificial-intelligence/>
 - 25 Interpol World. (2019). Engaging co-creation for future security threats. https://ecuritydelta.nl/images/INTERPOL_World_2019_Brochure.pdf
 - 26 Fuster, G.G., (2020). Artificial intelligence and law enforcement, Impact on fundamental rights. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)
 - 27 Information Commissioners Office. <https://ico.org.uk/>
 - 28 Information Commissioner's Office. (2017). Big data, artificial intelligence, machine learning and data protection; <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>
 - 29 Law Council of Australia., (2019). Artificial Intelligence: Australia's ethics framework. Department of Industry, Innovation and Science.
 - 30 UK Government., (2020, September 16). Data Ethics Framework. Central Digital and Data Office. <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020>
 - 31 European Commission., (2018). Artificial Intelligence for Europe. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>
 - 32 Evas, T. (2020). European framework on ethical aspects of artificial intelligence, robotics and related technologies. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU\(2020\)654179_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU(2020)654179_EN.pdf)
 - 33 Government of Canada. (2022). Responsible use of artificial intelligence (AI). <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1>
 - 34 European Union., and OEDC. (2021). National strategies on artificial intelligence. AI Watch. h
 - 35 National Institute of Standards and Technology. (2021). Building Trust in AI and ML. <https://www.nist.gov/system/files/documents/2021/08/18/ai-rmf-rfi-0014-attachment2.pdf>
 - 36 European Commission High-Level Expert Group on Artificial Intelligence., (2019). Ethics Guidelines for Trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
 - 37 Ibid.
 - 38 European Commission. (2018). Artificial Intelligence for Europe. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>
 - 39 Fuster, G.G., (2020). Artificial intelligence and law enforcement, Impact on fundamental rights. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)
 - 40 Fuster, G.G., (2020). Artificial intelligence and law enforcement, Impact on fundamental rights. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)
 - 41 Fuster, G.G., (2020). Artificial intelligence and law enforcement, Impact on fundamental rights. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)
 - 42 Akhgar et al., (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6> for easier reference
 - 43 Council of Europe. CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment. <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>
 - 44 Law Commission of Ontario., AI, ADM and the Justice System. <https://www.lco-cdo.org/en/our-current-projects/ai-adm-and-the-justice-system/>
 - 45 Law Council of Australia. (2019). Artificial Intelligence: Australia's ethics framework. Department of Industry, Innovation and Science.
 - 46 US Department of Homeland Security (DHS). (2021). S&T artificial intelligence and machine learning strategic plan. https://www.dhs.gov/sites/default/files/publications/21_0730_st_ai_ml_strategic_plan_2021.pdf
 - 47 The Alan Turing Institute & UK AI Council. (2021). AI Ecosystem Survey: Informing the National AI Strategy. Summary Report. https://www.turing.ac.uk/sites/default/files/2021-09/ai-strategy-survey_results_020921.pdf
 - 48 Centre for Data Ethics and Innovation. (2020). CDEI AI Barometer. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894170/CDEI_AI_Barometer.pdf
 - 49 Europol. Eurojust (June 2019). Common Challenges in Combating Cybercrime. <https://www.europol.europa.eu/publications-events/publications/common-challenges-in-combating-cybercrime>
 - 50 Committee on standards in public life., (2020). Artificial Intelligence and Public Standards. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF
 - 51 Fuster, G.G., (2020). Artificial intelligence and law enforcement. Impact on fundamental rights. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)
 - 52 Committee on standards in public life., (2020). Artificial Intelligence and Public Standards. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF
 - 53 Stix, C., (2021). Actionable principles for artificial intelligence policy: Three pathways. *Science and Engineering Ethics*. 27(15), 1-15.
 - 54 Mantelero, A. (2020). Elaboration of the feasibility study. Council of Europe.
 - 55 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI ethics for law enforcement: A study into

requirements for responsible use of AI at the Dutch police. Delphi, 2, 179.

- 56 Parliamentary Secretariat for Financial Services. (2019). Malta Towards Trustworthy AI. Office of the Prime Minister. https://malta.ai/wp-content/uploads/2019/10/Malta_Towards_Ethical_and_Trustworthy_AI_vFINAL.pdf
- 57 College of Policing. (2020). Policing in England and Wales: Future Operating Environment 2040. https://paas-s3-broker-prod-lon-6453d964-1d1a-432a-9260-5e0ba7d2fc51.s3.eu-west-2.amazonaws.com/s3fs-public/2020-08/Future-Operating-Environment-2040_0.pdf
- 58 Europol. (2019). Trustworthy AI Requires Solid Cybersecurity. <https://www.europol.europa.eu/media-press/newsroom/news/trustworthy-ai-requires-solid-cybersecurity>
- 59 UNODC. (2011). Handbook on police accountability, oversight and integrity. https://www.unodc.org/pdf/criminal_justice/Handbook_on_police_Accountability_Oversight_and_Integrity.pdf
- 60 Police Scotland. (2017). Policing 2026: Our 10 year strategy for policing in Scotland. <https://www.scotland.police.uk/spa-media/jjkn4et/policing-2026-strategy.pdf?view=Standard>
- 61 Kearns, I., and Muir, R. (2019). Data-driven Policing and Public Value. The Police Foundation. https://www.police-foundation.org.uk/2017/wp-content/uploads/2010/10/data_driven_policing_final.pdf
- 62 Police Professional. (2017). Harnessing potential of AI on front line. <https://www.policeprofessional.com/news/harnessing-potential-of-ai-on-front-line-2/>
- 63 National Security Commission on Artificial Intelligence. (2021). Final Report. <https://www.nsc.ai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>
- 64 Fair Trials. Regulating Artificial Intelligence for Use in Criminal Justice Systems in the EU. <https://www.fairtrials.org/app/uploads/2022/01/Regulating-Artificial-Intelligence-for-Use-in-Criminal-Justice-Systems-Fair-Trials.pdf>
- 65 Laat, P.B. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? Philosophy and Technology, 34, 1135-1193.
- 66 Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33-44).
- 67 Partnership on AI. (n.d.). Explainable AI in Practice. <https://partnershiponai.org/workstream/explainable-ai-in-practice/>
- 68 Partnership on AI. (2021). About ML. <https://partnershiponai.org/workstream/about-ml/>
- 69 Cutler, A., Pribic, M., & Humphrey, L., (2019). Everyday Ethics for Artificial Intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
- 70 Samsung. (2022). AI Ethics: To develop the best products and services with human resources and technology. <https://www.samsung.com/uk/sustainability/digital-responsibility/ai-ethics/>
- 71 Microsoft. (2022). Responsible AI: We are committed to the advancement of AI driven by ethical principles that put people first. <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimar6>
- 72 Accenture. (2022). Artificial Intelligence: AI Ethics and Governance. <https://www.accenture.com/gb-en/services/applied-intelligence/ai-ethics-governance>
- 73 Wagner, B. (2018) Ethics as an escape from regulation. From “ethics-washing” to ethics-shopping? In: E. Bayamlioglu, I. Baraliuc, L. Janssens et al. (eds.) Being Profiled: Cogitas Ergo Sum 10 Years of ‘Profiling the European Citizen’ (p. 84–88). Amsterdam: Amsterdam University Press.
- 74 Cutler, A., Pribic, M., & Humphrey, L., (2019). Everyday ethics for artificial intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
- 75 Laat, P.B. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? Philosophy and Technology, 34, 1135-1193.
- 76 Edelman. (2019). Edelman AI Survey. https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf
- 77 Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>
- 78 Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments. AI Now. <https://ainowinstitute.org/aiareport2018.pdf>
- 79 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life. <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>
- 80 Cutler, A., Pribic, M., and Humphrey, L., (2019). Everyday ethics for artificial intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
- 81 Jordan, S., Fazelpour, S., Koshiyama, A., Kueper, J., DeChant, C., Leong, B., Marchant, G., & Shank, C. (2019). Creating a Tool to Reproducibly Estimate the Ethical Impact of Artificial Intelligence. Pulse. <https://aipulse.org/creating-a-tool-to-reproducibly-estimate-the-ethical-impact-of-artificial-intelligence/>
- 82 De Almeida, P.G.R., dos Santos, C.D., and Farias, J.S., (2021). Artificial intelligence regulation: A framework for governance. Ethics and Information Technology, 23(3), 505-525.
- 83 Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>
- 84 UK.GOV. (2019). Understanding artificial intelligence ethics and safety. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>
- 85 OEDC. (2022). Recommendation of the Council on Artificial Intelligence. <https://oecd.ai/en/ai-principles>
- 86 European Commission. (2019). High-Level Expert Group on Artificial Intelligence. <https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf>
- 87 Standards Australia. (2019). An artificial intelligence standards roadmap. <https://www.standards.org.au/getmedia/ede81912-55a2-4d8e-849f-9844993c3b9d/1515-An-Artificial-Intelligence-Standards-Roadmap12-02-2020.pdf.aspx>

- 88 OECD. (2022). Recommendation of the Council on Artificial Intelligence. https://oecd.ai/en/ai-principles_p_3
- 89 IEEE., (2019). Ethically aligned design. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
- 90 IEEE., (2019). Ethically aligned design. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf, p. 99
- 91 Center for Democracy and Technology. So, you want to build an algorithm. <https://www.cdt.info/ddtool/>
- 92 Neudert, L.M., & Howard, P.N. (2020). Four principles for integrating AI and good governance. Oxford Commission on AI and Good Governance.
- 93 Standards Australia. (2019). An artificial intelligence standards roadmap. <https://www.standards.org.au/getmedia/ede81912-55a2-4d8e-849f-9844993c3b9d/1515-An-Artificial-Intelligence-Standards-Roadmap12-02-2020.pdf.aspx>
- 94 Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 95 Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 1, 389-399.
- 96 Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *30 Minds and Machines*, 99(102), 99-120.
- 97 Beckley, A., & Kennedy, M. (2020). Ethics and police practice. In: P. Birch, M. Kennedy, and E. Kruger. (eds). *Australian Policing: Critical Issues in 21st Century Police Practice*. London: Routledge.
- 98 Hagendorff, T., (2020). The ethics of AI ethics: An evaluation of guidelines. *30 Minds and Machines*, 99(102), 99-120; p. 112
- 99 Mantelero, A., & Esposito, M.S. (2021). An evidence based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law and Security Review*, (41), 1-35.
- 100 Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G., & Yu, H. (2015). Accountable Algorithms. *University of Pennsylvania Law Review*, 3(165), 633-707.
- 101 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper. https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf?sequence=3
- 102 Stix, C. (2021). Actionable principles for artificial intelligence policy: Three pathways. *Science and Engineering Ethics*, 27(15), 1-15.
- 103 Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of Explainability. *Science and Engineering Ethics*, 26(4), 2051-2068.
- 104 De Almeida, P.G.R., dos Santos, C.D., & Farias, J.S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, 23(3), 505-525.
- 105 Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of Explainability. *Science and Engineering Ethics*, 26(4), 2051-2068.
- 106 Wilson, C., Dalins, J., & Rolan, G. (2020). Effective, explainable and ethical: AI for law enforcement and community safety. *International Conference on Artificial Intelligence for Good (AI4G)*.
- 107 Schrader, D., & Ghosh, D. (2018). Proactively protecting against the singularity: Ethical decision making AI. *IEEE Computer and Reliability Societies Review*, 16(3), 56-63
- 108 De Almeida, P.G.R., dos Santos, C.D., & Farias, J.S., (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, 23(3), 505-525.
- 109 Mantelero, A., & Esposito, M.S. (2021). An evidence based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law and Security Review*, (41), 1-35.
- 110 Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M. F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. *NYU School of Law, Public Law Research Paper*, (20-54).
- 111 European Parliament. (2020). Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights. European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295\(SUM01\)_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295(SUM01)_EN.pdf)
- 112 European Commission High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 113 The Committee on Standards in Public Life., (2020). AI and Public Standards. A Review by the Committee on Standards in Public Life. <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>
- 114 Australian Government, (2019). Australia's Ethics Framework. A Discussion Paper. Department of Industry, Innovation and Science. <https://www.csiro.au/-/media/D61/Reports/Artificial-Intelligence-ethics-framework.pdf>
- 115 National Security Commission on Artificial Intelligence. (2021). Final Report. <https://www.nsc.ai.gov/2021-final-report/>
- 116 European Parliament. (2020). European framework on ethical aspects of artificial intelligence, robotics and related technologies. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU\(2020\)654179_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU(2020)654179_EN.pdf) ; European Parliamentary Research Service & European Parliament. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf)
- 117 Cutler, A., Pribic, M., & Humphrey, L. (2019). Everyday Ethics for Artificial Intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
- 118 FRA. (2020). Artificial Intelligence and Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf
- 119 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic accountability for the public sector. <https://www.opengovpartnership.org/documents/algorithmic->

[accountability-public-sector](#)

- 120 Leslie, D. (2019). Understanding AI Ethics and Safety: A guide for the responsible implementation of AI systems in the public sector. <https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety>
- 121 Amnesty International. (2019). PHRP Expert Meeting on Predictive Policing – Executive Summary. <https://www.amnesty.nl/content/uploads/2019/08/Expert-meeting-predictive-policing-executive-summary.pdf?x96671>
- 122 Neudert, N. D., & Howard, P. (2020). Four Principles for Integrating AI and Good Governance. Working paper 2020.1 Oxford, UK: Oxford Commission on AI & Good Governance
- 123 European Parliament. (2019). EU guidelines on ethics in artificial intelligence: Context and implementation. European Parliamentary. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf); Research Service & Council of Europe. (2020). AD Hoc Committee on Artificial Intelligence (CAHAI): Feasibility Study. <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>
- 124 UK Government. (2020, September 16). Data Ethics Framework. <https://www.gov.uk/government/publications/data-ethics-framework>; Government Digital Service & European Commission. (2021). 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 & Privacy International, Article 19., (2018). Privacy and Freedom of Expression in the Age of AI. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- 125 Interpol/UNICRI. (2020). Towards Responsible AI Innovation, Report on Artificial Intelligence for Law Enforcement. <https://ai-regulation.com/towards-responsible-ai-innovation-second-interpol-unicri-report-on-artificial-intelligence-for-law-enforcement>; Second Interpol-UNICRI Report on Artificial Intelligence for Law Enforcement & Interpol/UNICRI. (2019). Artificial Intelligence and Robotics for Law Enforcement. <http://www.unicri.it/artificial-intelligence-and-robotics-law-enforcement>
- 126 European Parliament. (2020). The Ethics of AI: Issues and Initiatives. European Parliamentary Research Service. [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU\(2020\)634452](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2020)634452)
- 127 Cutler, A., Pribic, M., and Humphrey, L., (2019). Everyday ethics for artificial intelligence. IBM. <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
- 128 Neudert, N. D., & Howard, P., (2020). Four Principles for Integrating AI and Good Governance. Working paper 2020.1 Oxford, UK: Oxford Commission on AI & Good Governance. <https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/124/2020/12/OxCAIGG-Report-Clibre-3.pdf>
- 129 Centre for Data Ethics and Innovation., (2020). Review into bias in algorithmic decision-making & OECD., (2021). Tools for Trustworthy AI. OECD Digital Economy Papers. OECD Publishing & European Union., (2019). Report with recommendations to the Commission on Civil Law Rules on Robotics. Committee on Legal Affairs.
- 130 Interpol., UNICRI. (2020). Towards Responsible AI Innovation, Report on Artificial Intelligence for Law Enforcement. Second Interpol-UNICRI Report on Artificial Intelligence For Law Enforcement & Interpol., UNICRI., (2019). Artificial Intelligence and Robotics for Law Enforcement
- 131 Gemeente Amsterdam, Helsinki, Saidot., (2020). Public AI Registers: Realising AI transparency and civic participation in government use of AI. <https://openresearch.amsterdam.nl/page/73074/public-ai-registers>
- 132 Council of Europe. (2020). Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary: Feasibility Study. European Commission for the Efficiency of Justice (CEPEJ). <https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4>
- 133 Interpol., UNICRI. (2020). Towards Responsible AI Innovation, Report on Artificial Intelligence for Law Enforcement and Interpol & UNICRI, 2019, Second Interpol-UNICRI Report on Artificial Intelligence for Law Enforcement. http://www.unicri.nu/in_focus/files/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_small.pdf
- 134 Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 135 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 136 Cavoukian, A., Taylor, S., & Abrams, M. E. (2010). Privacy by Design: essential for organizational accountability and strong business practices. Identity in the Information Society, 3(2), 405-413.
- 137 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>
- 138 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE.
- 139 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police. Delphi, 2, 179.
- 140 Centre for Data Ethics and Innovation. (2020). Review into Bias in Algorithmic Decision-Making.
- 141 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 142 Council of Europe. (2020). Ad Hoc Committee on Artificial Intelligence (CAHAI). Feasibility Study. CAHAI(2020)23.
- 143 European Commission. (2021). Proposal for a 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. European Commission. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- 144 Central Digital and Data Office & the Office for Artificial Intelligence. (2021). Ethics, Transparency and Accountability Framework for Automated Decision-Making. GOV.UK. <https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making#ensure-that-you-are-compliant-with-the-law>
- 145 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic

- Accountability for the Public Sector. Available at: [https:// www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/](https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/)
- 146 European Parliament. (2017). Report with Recommendations to the Commission on Civil Law Rules on Robotics. 2015/2103(INL).
 - 147 Information Commissioner's Office. (2017). Big data, artificial intelligence, machine learning and data protection. UK. <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>
 - 148 Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>
 - 149 High-Level Expert Group on Artificial Intelligence. (2019). Policy and Investment Recommendations for Trustworthy AI. European Commission.
 - 150 Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety>
 - 151 AI Now. (2018). Algorithmic Accountability Policy Toolkit. <https://ainowinstitute.org/aap-toolkit.pdf>
 - 152 Babuta, A. and Oswald, M. (2020). Data Analytics and Algorithms in Policing in England and Wales: Towards a New Policy Framework. RUSI.
 - 153 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police. Delphi, 2, 179.
 - 154 Haataja, M., van de Fliert, L., & Rautio, P. (2020). Public AI Registers: Realising AI Transparency and Civic Participation in Government Use of AI.
 - 155 Fundamental Rights Agency. (2020). Getting the Future Right - Artificial Intelligence and Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf
 - 156 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/>
 - 157 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE.
 - 158 Council of Europe. (2020). Ad Hoc Committee on Artificial Intelligence (CAHAI). Feasibility Study. CAHAI(2020)23.
 - 159 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
 - 160 Reed, C. (2018). How Should We Regulate Artificial Intelligence? <https://doi.org/10.1098/rsta.2017.0360>
 - 161 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police. Delphi, 2, 179.
 - 162 Holder, C., Khurana, V., & Mark, W. (2018). Artificial Intelligence: Public Perception Attitude and Trust. Technical Report, Bristow, 1-23.
 - 163 Center for Democracy & Technology. (n.d.). AI & Machine Learning. <https://cdt.org/ai-machine-learning/>
 - 164 Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 33(3), 371-377.
 - 165 Malgieri, G., & Comandé, G. (2017). Why a right to legibility of automated decision-making exists in the general data protection regulation. International Data Privacy Law.
 - 166 Holder, C., Khurana, V., & Mark, W. (2018). Artificial Intelligence: Public Perception Attitude and Trust. Technical Report, Bristow, 1-23.
 - 167 Select Committee on Artificial Intelligence. (2020). AI in the UK: No Room for Complacency. Report HL196, 2019-21. By the authority of the House of Lords.
 - 168 NPCC. & Association of Police and Crime Commissioners. (2020). National Policing Digital Strategy: Digital, Data and Technology Strategy 2020-2030. <https://www.apccs.police.uk/media/4886/national-policing-digital-strategy-2020-2030.pdf>
 - 169 European Commission. (2018). 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. Artificial Intelligence for Europe. Brussels, 25.4.2018 COM(2018) 237 final.
 - 170 Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M. F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. NYU School of Law, Public Law Research Paper, (20-54).
 - 171 Information Commissioner's Office. (2017). Big data, artificial intelligence, machine learning and data protection. UK.
 - 172 Select Committee on Artificial Intelligence. (2018). AI in the UK: Ready, Willing and Able? Report HL100, 2017-19. By the authority of the House of Lords.
 - 173 Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (18 April 2018). Algorithmic Accountability: A Primer. Data & Society. <https://datasociety.net/library/algorithmic-accountability-a-primer/>
 - 174 Law Council of Australia. (2019). Artificial Intelligence: Australia's Ethics Framework.
 - 175 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police. Delphi, 2, 179.
 - 176 BritainThinks. (2021). Complete Transparency, Complete Simplicity. Centre for Data Ethics and Innovation, and Central Digital and Data Office. GOV.UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995014/Complete_transparency_complete_simplicity_-_Accessible.pdf
 - 177 Ibid
 - 178 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/>
 - 179 Ibid
 - 180 Interpol & UNICRI. (2019). Artificial Intelligence and Robotics for Law Enforcement.

- 181 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/>
- 182 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.
- 183 Fundamental Rights Agency. (2020). Getting the Future Right - Artificial Intelligence and Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf
- 184 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 185 Ibid
- 186 Zardiashvili, L., Bieger, J., Dechesne, F., & Dignum, V. (2019). AI Ethics for Law Enforcement: A Study into Requirements for Responsible Use of AI at the Dutch Police. Delphi.
- 187 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 188 Ibid
- 189 Law Council of Australia. (2019). Artificial Intelligence: Australia's Ethics Framework.
- 190 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 191 Council of Europe. (2020). Ad Hoc Committee on Artificial Intelligence (CAHAI). Feasibility Study. CAHAI(2020)23.
- 192 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
- 193 High-Level Expert Group on Artificial Intelligence. (2019). Policy and Investment Recommendations for Trustworthy AI. European Commission.
- 194 Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & technology*, 31(4), 543-556.
- 195 Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). Four principles of explainable artificial intelligence. Gaithersburg, Maryland.
- 196 The Alan Turing Institute & UK AI Council. (2021). AI Ecosystem Survey: Informing the National AI Strategy. Summary Report.
- 197 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.
- 198 Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). Four principles of explainable artificial intelligence. Gaithersburg, Maryland.
- 199 Deloitte. (2021). Urban Future with a Purpose: 12 trends shaping the future of cities by 2030. Deloitte. <https://www2.deloitte.com/global/en/pages/public-sector/articles/urban-future-with-a-purpose/surveillance-and-predictive-policing-through-ai.html>
- 200 AI Now. (2018). Algorithmic Accountability Policy Toolkit. <https://ainowinstitute.org/aap-toolkit.pdf>
- 201 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE.
- 202 European Commission. (2018). 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. Artificial Intelligence for Europe. Brussels, 25.4.2018 COM(2018) 237 final.
- 203 Center for Democracy & Technology. (n.d.). AI & Machine Learning. <https://cdt.org/ai-machine-learning/>
- 204 Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H. V., Unsworth, K., Sahuguet, A., Venkatasubramanian, S., Wilson, C., Yu, C., & Zevenbergen, B. (n.d.). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. Fairness, Accountability, and Transparency. in *Machine Learning (FAT/ML)*. <https://www.fatml.org/resources/principles-for-accountable-algorithms>
- 205 Association for Computing Machinery US Public Policy Council (USACM). (2017). Statement on Algorithmic Transparency and Accountability. https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
- 206 Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). Four principles of explainable artificial intelligence. Gaithersburg, Maryland.
- 207 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/>
- 208 Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44).
- 209 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.
- 210 Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & technology*, 31(4), 543-556.
- 211 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper.
- 212 Partnership on AI. (n.d.). Explainable AI in Practice. <https://partnershiponai.org/workstream/explainable-ai-in-practice/>
- 213 Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., & Hajkowicz, S. (2019). Artificial intelligence: Australia's ethics framework-a discussion paper.
- 214 Pichai, S. (2018). AI at Google: our principles. *The Keyword*, 7, 1-3.
- 215 GCHQ. (2021). Pioneering a New National Security: The Ethics of Artificial Intelligence.
- 216 Council of Europe. (2001). Recommendation Rec (2001)10 adopted by the Committee of Ministers of the

- Council of Europe on 19 September 2001, p. 18.
- 217 European Parliament. (2020). Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights. European Parliament.
 - 218 Select Committee on Artificial Intelligence. (2018). AI in the UK: Ready, Willing and Able? Report HL100, 2017-19. By the authority of the House of Lords.
 - 219 The Committee on Standards in Public Life. (2020). Artificial Intelligence and Public Standards: A Review by the Committee on Standards in Public Life.
 - 220 Interpol & UNICRI. (2019). Artificial Intelligence and Robotics for Law Enforcement.
 - 221 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE.
 - 222 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic Accountability for the Public Sector. <https://www.opengovpartnership.org/documents/>
 - 223 Centre for Data Ethics and Innovation. (2020). Review into Bias in Algorithmic Decision-Making.
 - 224 OECD. (2021). Tools for Trustworthy AI: A Framework to Compare Implementation Tools for Trustworthy AI Systems. OECD Digital Economy Papers, No. 312. OECD Publishing.
 - 225 Ibid
 - 226 Regulation 2016/679 "General Data Protection Data Regulation" Article 14(2)(g) <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>
 - 227 Automated Decision Systems Accountability Act of 2021 CA A 13
 - 228 Algorithmic Accountability Act of 2019
 - 229 European Union (2021) Regulation on the European Parliament and of the Council – Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels, European Commission. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=SWD:2021:0085:FIN:EN:PDF>
 - 230 Council of the European Union. (2020). Presidency Conclusions on the Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change. <https://www.consilium.europa.eu/media/46496/st11481-en20.pdf>
 - 231 The Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA mirroring the GDPR in the law enforcement context.
 - 232 Regulation 2016/679 "General Data Protection Data Regulation" Article 14(2)(g)
 - 233 Automated Decision Systems Accountability Act of 2021 CA A 13
 - 234 Algorithmic Accountability Act of 2019.
 - 235 Kristen, K (2021) AI and Tort Law, In: F. Martin-Bariteau, T. Scassa (eds.), Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada). <https://ssrn.com/abstract=373465>
 - 236 Law Council of Australia. (2019). Artificial Intelligence: Australia's Ethics Framework' Department of Innovation and Science. <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework>
 - 237 HM Government. (2021). National AI Strategy. <https://www.gov.uk/government/publications/national-ai-strategy>
 - 238 HM Government. (2021). National AI Strategy. <https://www.gov.uk/government/publications/national-ai-strategy>
 - 239 HM Government. (2021). Algorithmic Transparency Standard. <https://www.gov.uk/government/collections/algorithmic-transparency-standard>
 - 240 Ibid
 - 241 HM Government. (2019). Artificial Intelligence and Public Standards – Written Evidence. <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-written-evidence>
 - 242 Hacker, P., Krestel, R., Grundmann, S. et al. (2020). Explainable AI under contract and tort law. Legal incentives and Technical Challenges. Artificial Intelligence Law, 28, 415-439.
 - 243 Council of Bars and Law Societies of Europe. (2020). CCBE Response to the consultation on the European Commission's White Paper on Artificial Intelligence. https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/IT_LAW/ITL_Position_papers/EN_ITL_20200605_CCBE-Response-to-the-consultation-regarding-the-European-Commission-s-White-Paper-on-AI.pdf
 - 244 EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) 22. https://edpb.europa.eu/our-work-tools/our-documents/edpb-edps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en
 - 245 For example, under Article 3 ECHR
 - 246 Fundamental Rights Agency. (2019). Facial recognition technology: fundamental rights considerations in the context of law enforcement. <https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law>
 - 247 Fundamental Rights Agency. (2019). Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights. <https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect>
 - 248 Fundamental Rights Agency. (2018). #BigData: Discrimination in data-supported decision making. <https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making>
 - 249 Fundamental Rights Agency. (2018). Preventing unlawful profiling today and in the future: A guide. <https://fra.europa.eu/en/publication/2018/preventing-unlawful-profiling-today-and-future-guide>
 - 250 Fundamental Rights Agency. (2020). Getting the Future Right - Artificial Intelligence and Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

- 251 Ibid
- 252 Ibid
- 253 Flexer, A., Dörer, M., Schlüter, J., Grill, T. (2018). Hubness as a case of technical algorithmic bias in music recommendation. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1062-1069.
- 254 Proposal for a 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, COM/2021/206 final
- 255 Access Now. (2021). An EU Artificial Intelligence Act for Fundamental Rights - A Civil Society Statement. <https://www.accessnow.org/cms/assets/uploads/2021/11/joint-statement-EU-AIA.pdf>
- 256 See also the concept of Materiality Threshold [cp. section on AP4AI Framework Blueprint](#)
- 257 Access Now. (2021). Here's how to fix the EU's Artificial Intelligence Act. <https://www.accessnow.org/how-to-fix-eu-artificial-intelligence-act>
- 258 European Digital Rights & Others (2021). European Digital Rights to Margrethe Vestager & Others
- 259 Access Now. (2021). Here's how to fix the EU's Artificial Intelligence Act. <https://www.accessnow.org/how-to-fix-eu-artificial-intelligence-act>
- 260 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 261 Ibid
- 262 Ibid
- 263 Council of Europe. (2020). Ad hoc Committee on Artificial Intelligence, 'Feasibility Study', Strasbourg, 17 December 2020 44, 50; Fundamental Rights Agency. (2020). Getting the Future Right - Artificial Intelligence and Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf, p. 87-98.
- 264 United Nations. (2011). Guiding Principles on Business and Human Rights. https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf
- 265 United Nations. (2011). Guiding Principles on Business and Human Rights. https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf
- 266 Data & Society. (2021). Mandating Human Rights Impacts Assessments in the AI Act. <https://ecnl.org/sites/default/files/2021-11/HRIA%20paper%20ECNL%20and%20Data%20Society.pdf>
- 267 Mantelero, A. (2020). Regulating AI within the Human Rights Framework: A Roadmapping Methodology, in: Phillip Czech et al. (eds), European Yearbook on Human Rights (p. 477-502). Cambridge University Press.
- 268 There is also a developing argument that Human Rights (like Data Protection) only cover the living whereas issues such as dignity and respect for the bodies of the dead and also genealogical considerations in DNA processing are a legitimate consideration in the LEA application of AI; see Response of the Biometrics & Surveillance Camera Commissioner to the consultation on data reform 2021 <https://www.gov.uk/government/publications/data-a-new-direction-commissioners-response/dcms-consultation-data-a-new-direction-response-by-the-biometrics-and-surveillance-camera-commissioner-accessible-version>
- 269 IEEE. (2019). Ethically aligned design. <https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e.pdf>
- 270 Alan Turing Institute. (2021). AI strategy survey results. https://www.turing.ac.uk/sites/default/files/2021-09/ai-strategy-survey_results_020921.pdf
- 271 Committee of Standards in Public Life. (2020). Artificial Intelligence and Public Standards. A Review by the Committee on Standards in Public Life. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF
- 272 OECD. (2021). State of implementation of the OECD AI Principles: Insights from national AI policies. OECD Digital Economy Papers, No. 311, OECD Publishing, Paris. <https://www.oecd.org/innovation/state-of-implementation-of-the-oecd-ai-principles-1cd40c44-en.htm>
- 273 Ibid
- 274 High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 275 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 276 Ada Lovelace Institute, AI Now Institute and Open Government Partnership. (2021). Algorithmic accountability for the public sector. <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector>
- 277 Gemeente Amsterdam, Helsinki, Saidot., (2020). Public AI Registers: Realising AI transparency and civic participation in government use of AI. <https://openresearch.amsterdam/nl/page/73074/public-ai-registers>
- 278 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 279 The recruitment across all countries was conducted by panel provider Qualtrics.
- 280 People, who failed to agree to one or more questions, were not allowed to proceed to ensure that all participants included in the consultation had given their full consent.
- 281 Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 282 Akhgar, B. (2003). Development of a methodology for design and implementation of SIS. SG publication, PEGAH 2003.
- 283 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 284 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>
- 285 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 286 A reference architecture often refer to a document or set of documents that provide overall structure of an implementation for a particular domain, it contain template solution which its instantiation can be utilised for customisation / re-use by maintaining consistency and applicability
- 287 See Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>

- 288 Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 289 Materiality is an assessment of the relative impact that something may have on accountability within the AI programme and in the context of Legality mirrors both the de minimis principle and the concept of proportionality.
- 290 Following the MoSCoW prioritisation approach, Must Have, Should Have, Could Have, Won't Have this time. It addresses the problems associated with simpler prioritisation approaches which are based on relative priorities such as risk. "The use of a simple high, medium or low classification is weaker because definitions of these priorities are missing or need to be defined. Nor does this categorization provide the business with a clear promise of what to expect. A categorisation with a single middle option, such as medium, also allows for indecision". The latter is particularly problematic in context of accountability towards AI deployment. The specific use of Must Have, Should Have, Could Have or Won't Have this time provides a clear indication of the weight of applicability of specific accountability principle for AI utilisation. See https://www.agilebusiness.org/page/ProjectFramework_10_MoSCoWPrioritisation
- 291 RACI is an acronym that stands for responsible, accountable, consulted and informed
- 292 In the next iterations of this report, we will develop and validate a guide and companion software tool to support organisations to create an AAA, rooted in applications of AI in the internal security domain. The guide will contain use cases and reference models for the application of different AI functions and domains based on the AP4AI principles.
- 293 For example, if an LEA adapt the MLOps model it involves the following steps: (a) Framing the business objectives (CONTEX), (b) Searching for the relevant data (SCOPE), (c) Preparing and processing the data (Data Engineering), (d) Developing and training the Machine Learning model, (e) Building and automating a machine learning pipeline (METHODOLOGY), (f) Deploy the model via static or dynamic deployment (METHODOLOGY) and Governance, compliance and security (GOVERNANCE). The mapping exercise should be carried out by the organisation to ensure the 12 Accountability Principles are applied in each stage of development life cycle.
- 294 A RACI chart or a responsibility assignment matrix— provide roles and responsibilities used in project management. A RACI chart defines whether the people involved in a project activity will be Responsible, 'Accountable' (at a tactical, task level), Consulted, or Informed for the corresponding task, milestone, or decision. <https://www.teamgantt.com/blog/raci-chart-definition-tips-and-example>.
- 295 Google has used the term "MLOps" where constantly monitoring the dataset and model are done all the way through design to deployment: <https://cloud.google.com/architecture/ml-ops-continuous-delivery-and-automation-pipelines-in-machine-learning>
- 296 Akhgar et al. (2022). AP4AI Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 297 These are examples of applicable laws, directives, procedures and rules; in the forthcoming report we will provide an extensive set of applicable laws for each principle.
- 298 For example, public procurement guidance
- 299 See e.g., The Danish Institute for Human Rights (2020) Guidance and Toolbox. https://www.humanrights.dk/sites/humanrights.dk/files/media/dokumenter/udgivelser/hria_toolbox_2020/eng/dihr_hria_guidance_and_toolbox_2020_eng.pdf.
- 300 Mantelero, A., & Esposito, M.S. (2021). An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems. Computer Law & Security Review, 41, doi:10.1016/j.clsr.2021.105561, <https://www.sciencedirect.com/science/article/pii/S0267364921000340>.
- 301 United Nations. (2011). Guiding Principles on Business and Human Rights https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.
- 302 See e.g., IEEE. (2019). Ethically Aligned Design, First Edition. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined; Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. (2019). Ethics Guidelines for Trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 303 See e.g. the tools provided by the French, Spanish and Italian data protection authorities respectively. <https://www.cnil.fr/en/privacy-impact-assessment-pia>; <https://www.aepd.es/es/derechos-y-deberes-cumple-tus-deberes/medidas-de-cumplimiento/evaluaciones-de-impacto>; <https://www.garantepriacy.it/regolamentoue/DPIA>. As regards the EU context, see also Article 29 Data Protection Working Party (2017). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. wp248rev.01.
- 304 See EDPB (2020) Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf. See also, ICO. Data protection by design and default. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/>.
- 305 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.
- 306 See e.g., Data Justice Lab. (2021). Advancing Civic Participation in Algorithmic Decision-Making: A Guidebook for the Public Sector. https://datajusticelab.org/wp-content/uploads/2021/06/PublicSectorToolkit_english.pdf
- 307 See Article 35.9 GDPR ("Where appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations").
- 308 See article 27, Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.
- 309 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.

- 310 See Council of Europe (2017) Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data, il of Europe 2017, Section IV, para 3.3. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a>
- 311 See e.g., Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108) (2019) Guidelines on Artificial Intelligence and Data Protection, T-PD(2019)01. <https://unesdoc.unesco.org/ark:/48223/pf0000377881>. (“AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based and ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially important role in areas where transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection”).
- 312 Independence is a requirement of national supervisory authorities, but not of the HRIA or DPIA, which in many cases are forms of self-assessment
- 313 Mantelero, A., & Esposito, M.S. (2021). An evidence based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law and Security Review*, (41), 1-35.
- 314 Cp. article 47 of the Charter of Fundamental Rights: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12016P047&from=EN>
- 315 It should be noted that enforcement relates to the activity of DPAs and is not part of the DPIA; on the other hand, the HRIA is not necessarily enforceable.
- 316 This is a general principle relating to the functioning of an organisation that is also reflected in its HRIA and DPIA practices, but not specifically addressed by them. The potential link relates to the required expertise and competence of those making the assessment.
- 317 Justice and Home Affairs
- 318 Europol's EU Terrorism Situation and Trend report (TE-SAT), published 23 June 2020
- 319 We refer to 2019 figures here as explained in Europol's 2021 Terrorism Situation and Trend Report (TESAT) it is not yet clear whether changes in 2020 are simply an artefact of the pandemic and the impact of the United Kingdom leaving the European Union. https://www.europol.europa.eu/cms/sites/default/files/documents/tesat_2021_0.pdf
- 320 Europol's Exploiting Isolation: Offenders and victims of online child sexual abuse during the COVID-19 pandemic, published 19 June 2020.
- 321 See e.g., Leclerc, B., Cale, J., Holt, T., and Drew, J. (2022). Child sexual abuse material online: The perspective of online investigators on training and support. *Policing: A Journal of Policy and Practice*, paac017, <https://doi.org/10.1093/police/paac017>
- 322 <https://www.microsoft.com/en-us/photodna>
- 323 National Center for Missing and Exploited Children. <https://www.missingkids.org/>
- 324 The text presented in [Appendix A](#) is taken from AP4AI Summary Report on Expert Consultations. <https://www.ap4ai.eu/node/6>
- 325 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.
- 326 Content reviews will be published in future iterations of this report.
- 327 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> v
- 328 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 329 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.
- 330 Jones, T., & Newburn, T. (2001). Widening Access: Improving Police Relations with Hard to Reach Groups. *Police Research Series Paper 138*, Home Office.
- 331 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.
- 332 In addition, COVID-19 pandemic restrictions hindered the AP4AI Project to meet experts in a face to face manner.

PROJECT COORDINATION

CENTRIC (Centre of Excellence for Terrorism, Resilience, Intelligence and Organised Crime Research): CENTRIC is a multi-disciplinary and end-user focused centre of excellence for end-user driven innovations in the field of security. The global reach of CENTRIC links both academic and professional expertise across a range of disciplines providing unique opportunities to progress groundbreaking research. The mission of CENTRIC is to provide a platform for researchers, practitioners, policy makers and the public to focus on applied security research. CENTRIC is a publicly funded organisation.

Europol: Europol is the European Union's Law Enforcement Agency. Europol hosts the EU Innovation Hub for Internal Security, a collaborative European network of innovation labs aimed at ensuring coordination and collaboration between EU internal security actors (law enforcement, justice, fundamental rights, border security, asylum, migration, customs etc.) in the field of innovation. The Hub supports the delivery of innovative solutions for internal security practitioners working for citizens' security in the area of freedom, security and justice. The EU Innovation Hub also contributes to establishing a common innovation picture for internal security and promote the alignment of innovation and security research efforts across the EU.

CONTACT

Accountability Principles for Artificial Intelligence (AP4AI)

Website: www.ap4ai.eu

Twitter: @AP4AI_project

Email: CENTRIC@shu.ac.uk; Innovation-lab@europol.europa.eu

