

ALIGNER D5.8

Final Update of the Research Roadmap for AI in
Support of Law Enforcement and Policing





Deliverable No.	D5.8
Work Package	WP5
Dissemination Level	PU
Author(s)	Daniel Lückerath, Valerie Wischott (Fraunhofer)
Co-Author(s)	Donatella Casaburo (KUL), Lindsay Clutterbuck (CBRNE), Mathilde Jarlsbo, Norea Normelli, Peter Svenmarck, Tommy Westmann (FOI)
Contributor(s)	Mattias Svahn (FOI)
Due date	2024-09-30
Actual submission date	2024-09-29
Status	Final
Revision	1.0
Reviewed by (if applicable)	Dominic Kelly (CBRNE), Anna Klose, Katharina Milde, Kai Pervölz (Fraunhofer), Christian Rabini (MPD), Adelina Zahirovic (SPA) Ari Basen (SIEAB), Shaban Buza (SIEAB), Penny Duquenoy (SIEAB), Philip Engström (SIEAB), Marco Filippi (SIEAB), Fredrik Heintz (SIEAB), Karl Hertting (LEAAB), Peter Kröjs (LEEAB), Andrius Paskauskas (SIEAB), Oliver Rose (SIEAB)

This document has been prepared in the framework of the European project ALIGNER – Artificial Intelligence Roadmap for Policing and Law Enforcement. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 101020574.

The sole responsibility for the content of this publication lies with the authors. It does not necessarily represent the opinion of the European Union. Neither the REA nor the European Commission are responsible for any use that may be made of the information contained therein.

Contact:

info@aligner-h2020.eu
www.aligner-h2020.eu



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement no. 101020574.



Executive Summary

The European Commission-funded Coordination and Support Action *ALIGNER: Artificial Intelligence Roadmap for Policing and Law Enforcement* brings together European actors concerned with Artificial Intelligence (AI), Law Enforcement, and Policing to collectively identify and discuss needs for paving the way for a more secure Europe in which AI supports police and law enforcement agencies (P&LEAs) while simultaneously empowering, benefiting, and protecting the public.

This deliverable presents the final iteration of the research roadmap, a key output not only of work package (WP) 5 “Outreach and Roadmap” but of the whole project. The roadmap compiles all the project results. Specifically, the roadmap

- ◆ presents the ALIGNER narrative – a vision of a potential future regarding the use of AI by criminals as well as P&LEAs;
- ◆ identifies practitioner needs that need to be met to counter (future) criminal use of AI and bring AI into service for P&LEAs;
- ◆ identifies and assesses AI technologies that can support practitioners under the postulated narrative;
- ◆ identifies how AI technologies might aid criminals in future and could lead to new crime patterns;
- ◆ identifies and discusses ethical, legal, societal, and organizational/technical implications of the use of AI by P&LEAs; and
- ◆ gives recommendations to policymakers and researchers on how to address the identified trends to meet the operational, cooperative, and collaborative needs of P&LEAs in the context of AI, while acknowledging ethical, legal, and societal implications.

To account for the broad network of actors in the fields of AI, law enforcement, and policing, ALIGNER’s research roadmap addresses

- ◆ LEA, policing, and criminal justice practitioners, including technical staff who are interested in applying, adapting, or co-creating upcoming research trends;
- ◆ research programmers and policymakers in local, regional, and national governments and other legislative bodies, who are interested in policy recommendations addressing identified gaps with regard to AI solutions for law enforcement;
- ◆ standardisation bodies to advance the unification of models, methods, tools, and data related to the use of AI in law enforcement;
- ◆ the research community surrounding AI, law enforcement and policing, as well as ethical, legal, and societal assessment; and
- ◆ the industry community surrounding AI and law enforcement who will receive directions for future developments and business opportunities.

The ALIGNER roadmap was iteratively developed, extended, and adapted over the course of three years, starting with the initial publication in September 2022, a second publication in March of 2023 and culminating in this final publication in September 2024. The majority of its content results from work conducted by individual project partners, three online surveys that ran between 2022 and 2024, eight workshops held by ALIGNER with practitioners from law enforcement and policing, research and academia, industry professionals, and policymakers between 2021 and 2024, as well as expert discussions during several research and policy events.



The work of ALIGNER – and subsequently this roadmap – assumes a vision of the future where AI is a constant criminal threat, and a regular tool used by law enforcement agencies. Within this vision, ALIGNER focuses on a limited number of topical areas with highest relevance for P&LEAs and other actors in the field of law enforcement and AI: disinformation and social manipulation, cybercrime against individuals and organizations, and support of policing on the city-level. These high-interest topics are captured in an overall narrative: a high-level description of a potential (near-term) future, including how AI might be used for criminal behaviour as well as to support P&LEAs.

Based on this narrative ALIGNER has identified capability enhancement needs of P&LEAs, potential criminal misuses of AI technologies, as well as ethical and legal issues related to the use of AI by P&LEAs. These findings are addressed by identifying and assessing a set of AI technologies for their potential to enhance the capabilities of P&LEAs, the associated risks of using these technologies, and potential mitigation measures.

Based on these findings, the project worked with practitioners from P&LEAs, research, policymaking, industry, and civil society, to develop nine policy recommendations and 19 research recommendations. The research recommendations comprise recommendations for research into ethical, legal, and societal implications of the (mis)use of AI, as well as technical issues necessary to enable better and more secure use of AI. The policy recommendations are:

1. Ensure a constructive partnership between the AI Office and Member States' P&LEAs to ensure prevention of compliance issues, identification and exchange of best practices / lessons learned, and facilitate the joint co-creation of targeted guidelines for the implementation of the AI Act at P&LEAs.
2. Explore the use of the EU Database for High-Risk AI systems by P&LEAs to facilitate exchange between European P&LEAs about the development, deployment, and use of High-Risk AI systems in compliance with the AI Act.
3. Clarify the meaning of “*a genuine and present of foreseeable threat of a terrorist attack*” in Article 5 of the AI Act to ensure ethical and legal use of remote biometric identification by P&LEAs.
4. Embed the concepts of ‘AI Literacy’ and ‘human-centric approach’ into EU P&LEA training, including education on impacts, consequences, and implications of AI system use as well as use of real-world data for model training.
5. Establish and improve unified frameworks, compliant with the AI Act, for the evaluation of AI systems and models during development and deployment ensuring their ethical, legal, and societal compliance.
6. Review existing and establish new legal and regulatory mechanisms to ensure that AI systems and their use are ethical, legal, and societally acceptable.
7. Develop meaningful dialogue between regulators, P&LEAs, researchers, industry, and civil society organizations to strengthen citizens' confidence in the use of AI systems by P&LEAs via the consultation processes of the AI Office and other means.
8. Enable EU citizens to access basic information about AI systems used by P&LEAs.
9. Extend and adapt European and national research programmes to better facilitate evidence-based, participatory research into P&LEA needs regarding AI, the potential implications of the use of AI by P&LEA, and potential criminal use of AI.



Table of Contents

Executive Summary	3
List of Abbreviations	7
1. Introduction	8
1.1 What's New in This Version?	9
2. A Potential Vision of the Future	10
2.1 Background	10
2.2 The Narrative	10
3. Practitioner Capability Enhancement Needs	13
3.1 Status Quo of AI in Law Enforcement and Policing (2022-2023)	13
3.2 Potentials of AI in Law Enforcement and Policing	15
3.3 Perceived Challenges of AI in Law Enforcement and Policing	17
Perceived Challenges Related to the ALIGNER Narrative	18
3.4 Perceived Trends in the Criminal Misuse of AI	19
4. Implications of Broader Distribution of AI Technology	21
4.1 Ethical and Legal Aspects	21
4.2 Cybersecurity Aspects of AI	25
4.3 Potential Malicious Use of AI	26
4.3.1 Disinformation and Social Manipulation	26
4.3.2 Cybercrime Against Individuals and Organizations	28
4.3.3 Vehicles, Robots, and Drones	30
5. Policy and Research Recommendations	32
5.1 The Relevant Legal Framework: The AI Act	32
5.1.1 Scope of application	32
5.1.2 Risk-based approach	33
5.2 Policy Recommendations	38
5.2.1 Recommendation overview	39
5.2.2 Recommendations in detail	40
5.3 Research recommendations	48
5.3.1 Short- to medium-term research	48
5.3.2 Medium- to long-term research	52
6. AI Technology Catalogue	56
6.1 Deanonymization – Authorship Attribution	57
6.2 Deanonymization – Geolocalisation of Images	59
6.3 Veracity Assessment – Disinformation Detection	61
6.4 Detection of Synthetic Images	63



6.5	Detection of Synthetic Video	65
6.6	Language Models	67
6.7	Automatic Detection of Scammer Profiles	69
6.8	Automatic Identification of Potential Scam Victims	71
6.9	Detection of Voice Clones	73
6.10	Detection of Crypto Currency Laundering	75
6.11	Using Drones for Handling an IED Incident	77
6.12	Facial Recognition	79
6.13	Drone Use for Object Detection	81
6.14	Chatbots	83
7.	Conclusions	85
8.	References	86
	Annex A: Projects and Initiatives Mapping	89
	Annex B: Additional information on the online surveys	94
	First ALIGNER survey: 2022	94
	Demographic information on the survey sample	95
	Unprioritized and categorized answers on potentials and challenges	96
	Second ALIGNER survey: 2023	99
	Demographic information on the survey sample	99
	Unprioritized and categorized answers on examples for use of emerging AI technologies	101



List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AML	Anti-Money Laundering
API	Application Programming Interface
BSI	Bundesamt für Sicherheit in der Informationstechnik (German Federal Office for Information Security)
CCTV	Closed-Circuit Television
CEPOL	European Union Agency for Law Enforcement Training
DDoS	Distributed Denial-of-Service
ECHR	European Convention on Human Rights
ENISA	European Union Agency for Cybersecurity
EU	European Union
IED	Improvised Explosive Device
LLM	Large Language Model
ML	Machine Learning
N	Sample size
NLP	Natural Language Processing
P&LEA	Police & Law Enforcement Agency
SAR	Search & Rescue
TFEU	Treaty on the Functioning of the EU
TRL	Technology Readiness Level
UAV/S	Unmanned Aerial Vehicles / Systems
WP	Work Package



1. Introduction

This deliverable has been prepared for the European Commission-funded Coordination and Support Action *ALIGNER: Artificial Intelligence Roadmap for Policing and Law Enforcement*. ALIGNER aims to bring together European actors concerned with AI, law enforcement, and policing to collectively identify and discuss needs for paving the way for a more secure Europe in which AI supports P&LEAs while simultaneously empowering, benefiting, and protecting the public. To achieve this, ALIGNER

- (1) facilitated communication and cooperation between actors from law enforcement, policing, policymaking, research, industry, and civil society about the changing dynamics of crime patterns relevant to the use of AI by establishing a workshop series;
- (2) identified the capability enhancement needs of European LEAs;
- (3) identified, assessed, and validated AI technologies with potential for P&LEA capability enhancement by implementing a technology watch process that includes impact and risk assessments;
- (4) identified ethical, societal, and legal implications of the use of AI in law enforcement;
- (5) identified potential criminal uses of AI via the development of a taxonomy of AI-supported crime;
- (6) identified policy and research needs related to the use of AI in law enforcement; and
- (7) employed the gathered insights to incrementally develop and maintain this AI research roadmap.

This deliverable presents the final iteration of the research roadmap, a key output not only of WP 5 “Outreach and Roadmap” but of the whole project. The roadmap compiles all the project results. Specifically, the roadmap

- ◆ presents the ALIGNER narrative – a vision of a potential future regarding the use of AI by criminals as well as P&LEAs;
- ◆ identifies practitioner needs that need to be met to counter (future) criminal use of AI and bring AI into service for P&LEAs;
- ◆ identifies and assesses AI technologies that can support practitioners under the postulated narrative;
- ◆ identifies how AI technologies might aid criminals in future and could lead to new crime patterns;
- ◆ identifies and discusses ethical, legal, societal, and organizational/technical implications of the use of AI by P&LEAs; and
- ◆ gives recommendations to policymakers and researchers on how to address the identified trends to meet the operational, cooperative, and collaborative needs of P&LEAs in the context of AI, while acknowledging ethical, legal, and societal implications.

To account for the broad network of actors in the fields of AI, law enforcement, and policing, ALIGNER’s research roadmap addresses

- ◆ LEA, policing, and criminal justice practitioners, including technical staff who are interested in applying, adapting, or co-creating upcoming research trends;
- ◆ research programmers and policymakers in local, regional, and national governments and other legislative bodies, who are interested in policy recommendations addressing identified gaps with regard to AI solutions for law enforcement;
- ◆ standardisation bodies to advance the unification of models, methods, tools, and data related to the use of AI in law enforcement;



- ◆ the research community surrounding AI, law enforcement and policing, as well as ethical, legal, and societal assessment; and
- ◆ the industry community surrounding AI and law enforcement who will receive directions for future developments and business opportunities.

The content of the roadmap results from work conducted by individual project partners, three online surveys conducted between 2022 and 2024, as well as eight workshops held by ALIGNER with practitioners from law enforcement and policing, research and academia, industry professionals, and policymakers between 2021 and 2024. In addition, ALIGNER partners participated in expert discussions during several research and policy events. Lastly, ALIGNER intensively exchanged with its sibling projects popAI¹ and STARLIGHT² as well as the EU project AP4AI³, that together with ALIGNER form the AI cluster of EU research projects.

The roadmap is structured as follows: This section continues with a short description of what is newly included or modified in this iteration of the document. Section 2 then introduces the ALIGNER narrative, before section 3 gives an overview of the identified capability enhancement needs of law enforcement practitioners and further – positive and negative – trends and potentials of AI technologies. Section 4 then continues with an overview of the implications stemming from the broader use of AI technologies in society in general and in the context of law enforcement specifically, with a special focus on ethical and legal aspects, cybersecurity requirements, and the potential malicious use of AI. Section 5 starts with an overview of the most relevant EU policy regarding AI – the AI Act – before providing the main output of the roadmap: policy and research recommendations. The roadmap closes with the AI technology catalogue – a detailed overview of the AI technologies identified for the ALIGNER narrative, including an assessment for their ethical, legal, and technological risks and suggestions for potential mitigation measures. In addition, the annex to the roadmap provides an overview of relevant research projects in the field of AI and more detailed information from ALIGNER’s first and second online survey.⁴

1.1 What’s New in This Version?

This final version of the ALIGNER roadmap, submitted in September 2024, brings all results of the project together. It provides one coherent narrative for all three scenario topics considered in ALIGNER, discusses the identified capability enhancement needs of P&LEAs, including developments stemming from the large-scale, public release of generative AI models in late 2022, identifies potential AI misuse and cybersecurity issues, extends the policy recommendations, and provides suggestions for research directions. It also includes several revisions in all sections – in addition to a restructuring to enhance reading flow.

¹ H2020 popAI – A European Positive Sum Approach towards AI tools in support of Law Enforcement and safeguarding privacy and fundamental rights, 2021-2023: <https://www.pop-ai.eu/> [Accessed 2024-09-27]

² H2020 STARLIGHT - Sustainable Autonomy and Resilience for LEAs using AI against High priority Threats, 2021-2025: <https://starlight-h2020.eu/> [Accessed 2024-09-27]

³ AP4AI – Accountability Principles for AI: <https://www.ap4ai.eu/> [Accessed 2024-09-27]

⁴ Information on the third survey can be found in ALIGNER D3.3 [6].



2. A Potential Vision of the Future

2.1 Background

The work of ALIGNER – and subsequently this roadmap – assumes a vision of the future where AI is a constant criminal threat, and a regular tool used by P&LEAs. Within this vision, ALIGNER focuses on a limited number of topical areas with highest relevance for P&LEAs and other actors in the field of law enforcement and AI: disinformation and social manipulation, cybercrime against individuals and organizations, and support of policing on the city-level. These high-interest topics are captured in an overall narrative: a high-level description of a potential (near-term) future, including how AI might be used for criminal behaviour as well as to support P&LEAs.⁵ The focus of the narrative was selected based on expert input from ALIGNER’s advisory boards⁶ and in collaboration with several other research projects. The selection was then validated via an online survey that ran between May and August 2022 (see also Annex B).

2.2 The Narrative

AI has penetrated all aspects of daily life. Its applications span across smart phones, computers, (smart) home devices, personal assistants and decision-support systems. Sadly, AI is also exploited by individuals, organized crime syndicates, and state-sponsored malicious actors to commit crimes and carry out malicious acts. Meanwhile, P&LEAs utilize AI to prevent, detect, and counteract criminal activities, uncover patterns for further investigations, and to support their activities and operations.

Phishing attacks are one of the most damaging uses of AI by criminals, both against individuals and organisations. Traditional phishing entails sending deceptive communications that appear credible to steal sensitive information like passwords or credit card numbers. AI has made this more advanced through ‘spear phishing’ attacks. Using extensive data analysis, AI identifies valuable targets and creates highly personalized messages. Voice cloning of individuals by AI systems has opened new opportunities for targets to be deceived – even more so in combination with cloned images and videos. For instance, a criminal might use AI to scan social media profiles and email exchanges to compose a phishing email or voicemail that seems to come from a trusted colleague or friend, increasing the likelihood of obtaining sensitive information.

With the acquired sensitive information, criminals can deploy AI in sophisticated cybercrime activities. AI-driven malware evolves and adapts autonomously to bypass security measures, presenting a significant threat. These advanced programs use machine learning (ML) to find vulnerabilities within an organization’s network, reducing the effectiveness of traditional defence mechanisms. Additionally, AI-powered bots can execute distributed denial-of-service (DDoS) attacks, overwhelming systems with traffic and causing operational disruptions. These AI-backed cyber threats necessitate equally

⁵ In the working context of ALIGNER, the overarching vision of the future is also called the ‘archetypical scenario’. Within this vision, ALIGNER build several ‘scenarios’ that specified potential uses of AI by criminals as well as P&LEA and related implications. These scenarios have been fleshed out using a ‘narrative’. See also ALIGNER D2.2 [1] for additional details.

⁶ Over the runtime of the project, ALIGNER maintained two advisory boards, which brought together over 60 external experts from across the EU from law enforcement, research, ethics, law, industry, and policymaking. More on the advisory boards can be found in ALIGNER D2.1 [25]



advanced AI solutions for detection and mitigation, as conventional cybersecurity measures often fail against such dynamic and intelligent threats.

In exploiting individuals, romance scams have also become more complex, often involving cryptocurrency. Both solo operators and organized crime groups use AI to conduct 'crypto romance scams' on a large scale. The scammer builds a relationship with the victim through online interactions, gradually earning trust. Once trust is established, the scammer proposes investing in cryptocurrency, promising high returns. Convinced of the scammer's authenticity, the victim transfers money to a cryptocurrency wallet suggested by the scammer. The scheme may end abruptly with the scammer and funds disappearing, or it may continue, extracting more money over time. AI chatbots managing multiple conversations simultaneously often facilitate this manipulation, making the scam more scalable and effective.

On a societal level, AI's ability to create realistic videos, images, texts, and sounds has introduced a new kind of threat: disinformation and social manipulation. Malicious entities use AI to generate deepfake videos, pictures, and audio that realistically depict public figures making false statements or participating in events that never occurred. These manufactured pieces can be spread across social media to sway public opinion or incite unrest. For example, during an election cycle, criminals might release a deepfake video of a candidate making incendiary remarks, thereby influencing voters and undermining the democratic process. Another example would be the manipulation of images or videos of celebrities – or everyday citizens – to falsely suggest their support of a political candidate in order to sway public opinion.

To counter the threats posed by AI-enhanced disinformation, law enforcement agencies employ their own AI systems for veracity assessment and deanonymisation. These tools help detect disinformation by analysing patterns in data and cross-referencing with verified sources. Additionally, AI can attribute authorship and geolocate images, tracing the origin of false information. For example, when a deepfake video surfaces, AI can analyse the video's metadata and compare it against a database of known authentic videos, thereby determining its authenticity and potentially identifying the perpetrator.

AI also plays a crucial role in countering AI-enabled cybercrime by providing advanced detection and mitigation techniques that traditional methods cannot match. Through ML algorithms, AI can identify patterns and anomalies within vast datasets, flagging potential threats like malware and spear phishing attempts before they can cause harm. For instance, AI-driven systems can analyse email metadata and content to detect phishing scams by comparing them with known malicious signatures and suspicious behavioural patterns. Additionally, AI enhances digital forensics by swiftly sifting through enormous amounts of data to uncover hidden connections and potential vulnerabilities exploited by cybercriminals. In the realm of cryptocurrency scams, AI tools can trace transactions and identify fraudulent activities in real-time, providing law enforcement with actionable insights to intercept and prevent further crimes. Overall, AI's ability to process and analyse information at unprecedented speeds makes it an indispensable asset in the fight against sophisticated cyber threats.

AI also significantly enhances situational awareness and incident response at the city level for law enforcement. Information from multiple sources, such as public reports, police databases, and CCTV footage, is processed by AI to provide a comprehensive and coherent picture of an ongoing incident. For instance, during a large public event, AI can filter and manage incoming data, identify potential threats, and suggest optimal response strategies. This capability is particularly valuable in high-



pressure situations where rapid decision-making is critical. By correlating data from various inputs, AI helps incident commanders make informed decisions, ensuring a swift and effective response.

AI's ability to handle and process large volumes of data also makes it indispensable for law enforcement in other areas. For example, AI systems can analyse fingerprints, facial features, and other biometric data at a speed and accuracy unmatched by human operators. AI-driven surveillance systems can – under certain legal criteria – monitor public spaces, identifying suspicious activities and alerting authorities promptly. Furthermore, AI can assist in forensic investigations by analysing digital evidence, such as text messages and emails, to uncover hidden patterns or connections that would be difficult for humans to detect. These applications not only enhance the efficiency of P&LEAs but also significantly improve public safety.



3. Practitioner Capability Enhancement Needs

To identify in which areas of law enforcement and policing work AI can unfold the most potential and to identify potential barriers for the deployment of AI the ALIGNER team assessed

- ◆ the use of AI by P&LEAs;
- ◆ the areas in which practitioners, researchers, and other actors in the field of AI, law enforcement, and policing identify the highest potential of AI; and
- ◆ where they see the largest challenges when introducing AI.

This information was gathered during the ALIGNER workshops as well as via a series of online surveys between 2022 and 2024.⁷ The first survey was run between May and August 2022 – before the publication of ChatGPT by OpenAI – and established the then current practices with, potentials of, and challenges for the use of AI by P&LEAs. The second survey was run between May and August 2023 – after the publication of ChatGPT – and was conducted to understand how recent developments in the AI fields had impacted the work of P&LEAs. The third and final survey was run between March and May 2024 to receive an updated picture of the potentially malicious use of AI.

3.1 Status Quo of AI in Law Enforcement and Policing (2022-2023)⁸

When discussing the use of AI with P&LEAs, it becomes evident that at present, AI is not used at all or only to a limited extent in the operative work of most P&LEAs – even after the advent of Large Language Model (LLM)-powered generative AI systems like ChatGPT. However, usage seems to be increasing. This is supported by the survey results. In 2022, 17 P&LEA practitioners indicated that AI is currently used to a very little or some extent, compared against 21 practitioners who indicated use of AI at least to little extend in 2023 (see Figure 1).⁹ This is also indicated by

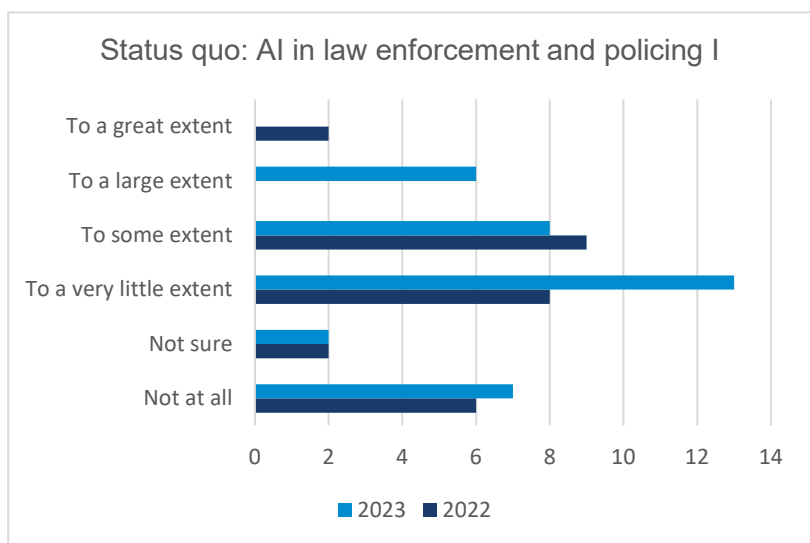


Figure 1: Results of the question “To what extent is AI currently being applied in your work?”, first and second ALIGNER survey, 2022/23.

the number of P&LEAs who use AI to a large or great extent (0 and 2 in 2022 vs. 6 and 0 in 2023). The number of P&LEA not making use of AI remained almost constant between 2022 and 2023 (6 vs. 7 responses). These results are not surprising as discussions with practitioners showed that many P&LEAs still grapple with the basic technological demands for the use of AI and the leadership in many P&LEAs needs to be convinced of the fundamental impact AI can and will have on their organisations

⁷ For more information on the first two surveys, please see ANNEX B. For additional information on the third survey, please see ALIGNER D3.3 [6].

⁸ Please note that the respondents of survey one and survey two not necessarily overlap. While both surveys were distributed using the same channels, the second survey was not only distributed to participants of the first survey.

⁹ This and the following question (Figure 2) should only be answered by P&LEA. However, the number of responses differed between those two questions. This means that the samples could also include non-practitioners.



to support broader use of AI. P&LEAs also indicated in interviews that usually only highly specialized cybercrime units currently employ AI to a great extent, as AI is a prerequisite for their daily work. In other P&LEA departments the use of AI is instead much more dependent on individual motivation of investigators, e.g. an investigator seeking additional specific capacities, or someone employed in a research department wanting to examine the use of a novel technology.

At the same time, many P&LEAs are convinced that AI can enhance existing functions and capabilities or enable the development of new capabilities.

However, the extent to which AI has brought benefits varies (Figure 2). Most respondents of ALIGNER's first survey in 2022 indicated that the functions and capabilities of P&LEAs have benefitted to some extent, with fewer respondents indicating that they have benefitted largely or to a great extent. However, no one indicated that functions and capabilities have not improved at all, but at least to a very little extent. From the survey sample, it appears that AI has enabled the development of new functions and capabilities rather than improving existing functions and capabilities. This is reinforced by results from ALIGNER's second survey, where roughly half of all P&LEA respondents (N=23) indicated that new AI applications or other new tools using AI were increasingly used by P&LEA (Figure 3, New tools I) and more than 75% of all P&LEA respondents (N=27) indicated that there are plans to make use of recently emergency AI applications, tools, and technologies (Figure 3, New tools II).

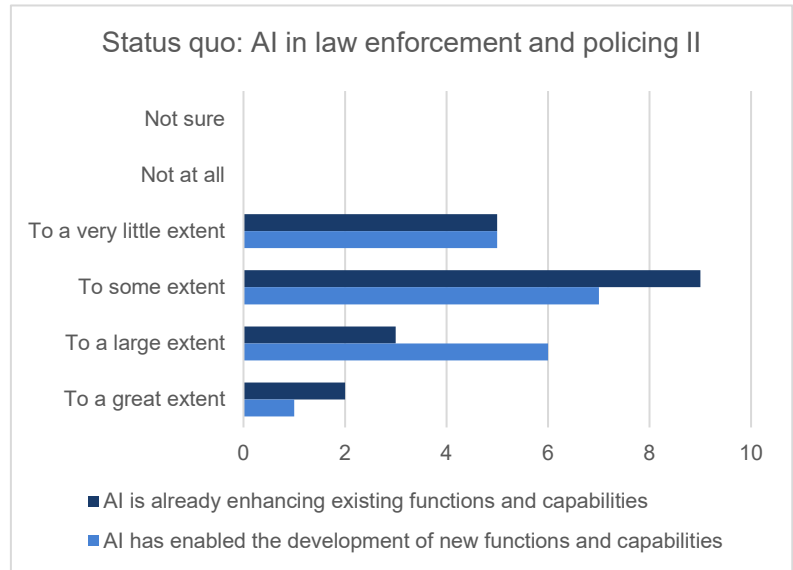


Figure 2: Results of the questions “To what extent do you think the use of AI is enhancing existing/has enabled the development of new functions and capabilities in law enforcement and policing?”, first ALIGNER survey, 2022.

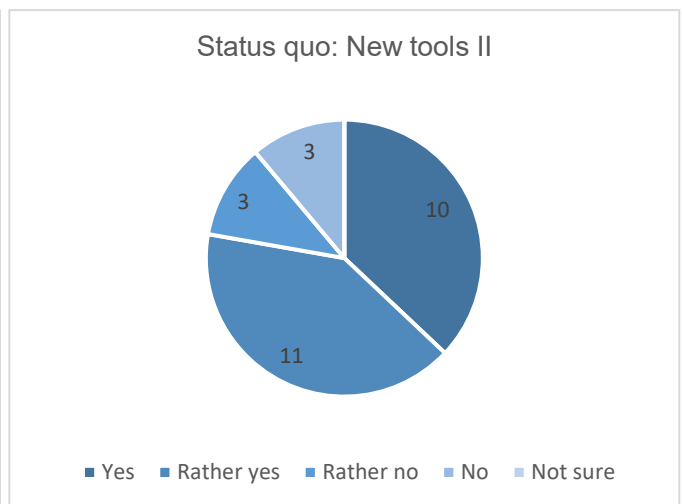
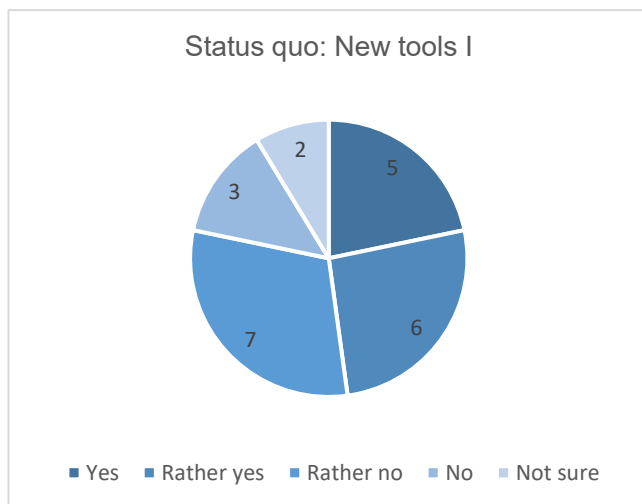


Figure 3: Results of the questions “With the rapid public emergence of new AI applications in recent months, are new AI applications or other new tools using AI technologies being increasingly used in policing and law enforcement?” (New tools I) and “Are you aware of plans to (increasingly) use the recently emerging AI applications, tools and technologies in the work of police and law enforcement agencies in the future?” (New tools II), second ALIGNER survey, 2023.



3.2 Potentials of AI in Law Enforcement and Policing

Considering that AI is only used to a limited extent by P&LEAs – although if used seems to enable the development of new capabilities – and is planned to be increasingly used, the question arises: In which areas of work would AI have the greatest impact? All participants of ALIGNER’s workshops hinted at the high relevance of AI for P&LEAs and the first survey results support this assessment (Figure 4). Indeed, 95% of the participants stated that it is “relevant” or “very relevant”¹⁰.

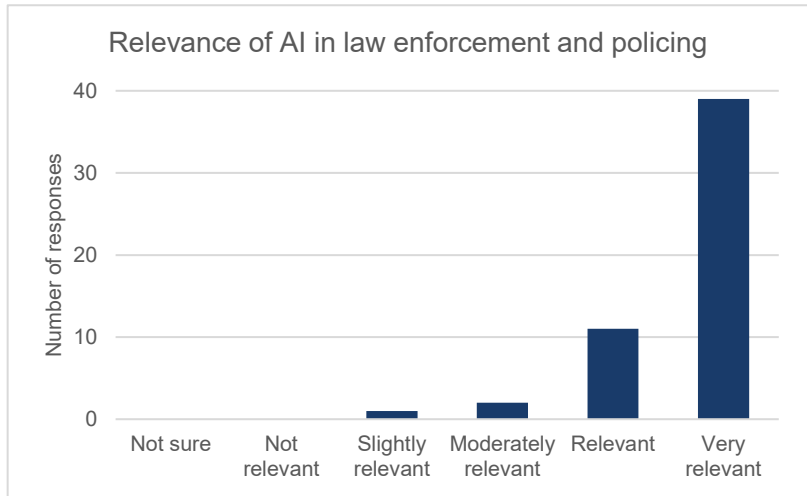


Figure 4: Results of the question “How relevant do you think the use of AI is in law enforcement and policing?”, first ALIGNER survey, 2022.

To identify specific work areas in which AI might support P&LEAs, ALIGNER delineated seven different categories of P&LEA capabilities and functions based on working sessions held during the first two ALIGNER workshops (see Figure 6 and ALIGNER D2.2 [1]). During the first two ALIGNER workshops, P&LEA practitioners as well as researchers and industry professionals, unsurprisingly, identified those work areas that are heavily dependent on data as most promising for the application of AI. The survey responses support these results: In 2022 participants were asked to rate the extent to which each of the named functions and capabilities could benefit from the use of AI (Figure 6). The highest level of agreement is found in data and information handling processes, where almost 90% of the 56 participants believe, they could benefit to a large or great extent from the use of AI. This is followed by biometric recognition and identification (83%¹¹), digital forensics (81%) and the detection and prevention of crimes and threats within the digital domain (78%). There is less consensus for incident reaction and response (65%), autonomous vehicles, robots, and drones (64%), and the detection and prevention of crimes and threats outside the digital domain (56%). Confirming these findings, the second ALIGNER survey showed that biometric recognition and identification as well as data and information handling processes are those areas in which P&LEA mostly apply AI (Figure 5)¹².

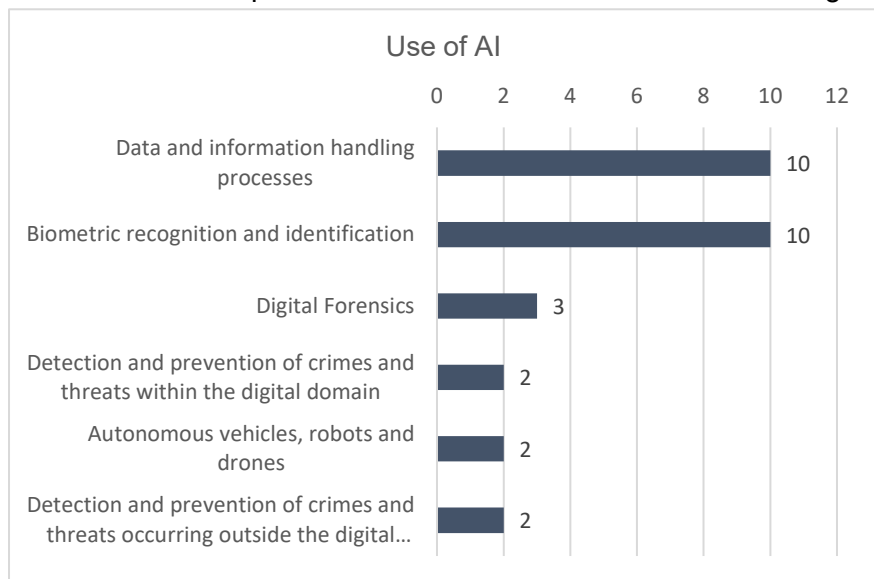


Figure 5: Results of the question "If you are aware of it, in which areas of policing and law enforcement are AI application applied", second ALIGNER survey, 2023.

¹⁰ This question and all the following questions in this section were answered by all participants.

¹¹ This and the following numbers in brackets refer to participants who answered with “to a large extent” and “to a great extent”.

¹² Please see Annex B for a list of examples for AI use, provided by participants of the second survey. Note that the category “incident reaction and response” is not included in Figure 5, as respondents of the second survey did not provide any examples for this category.

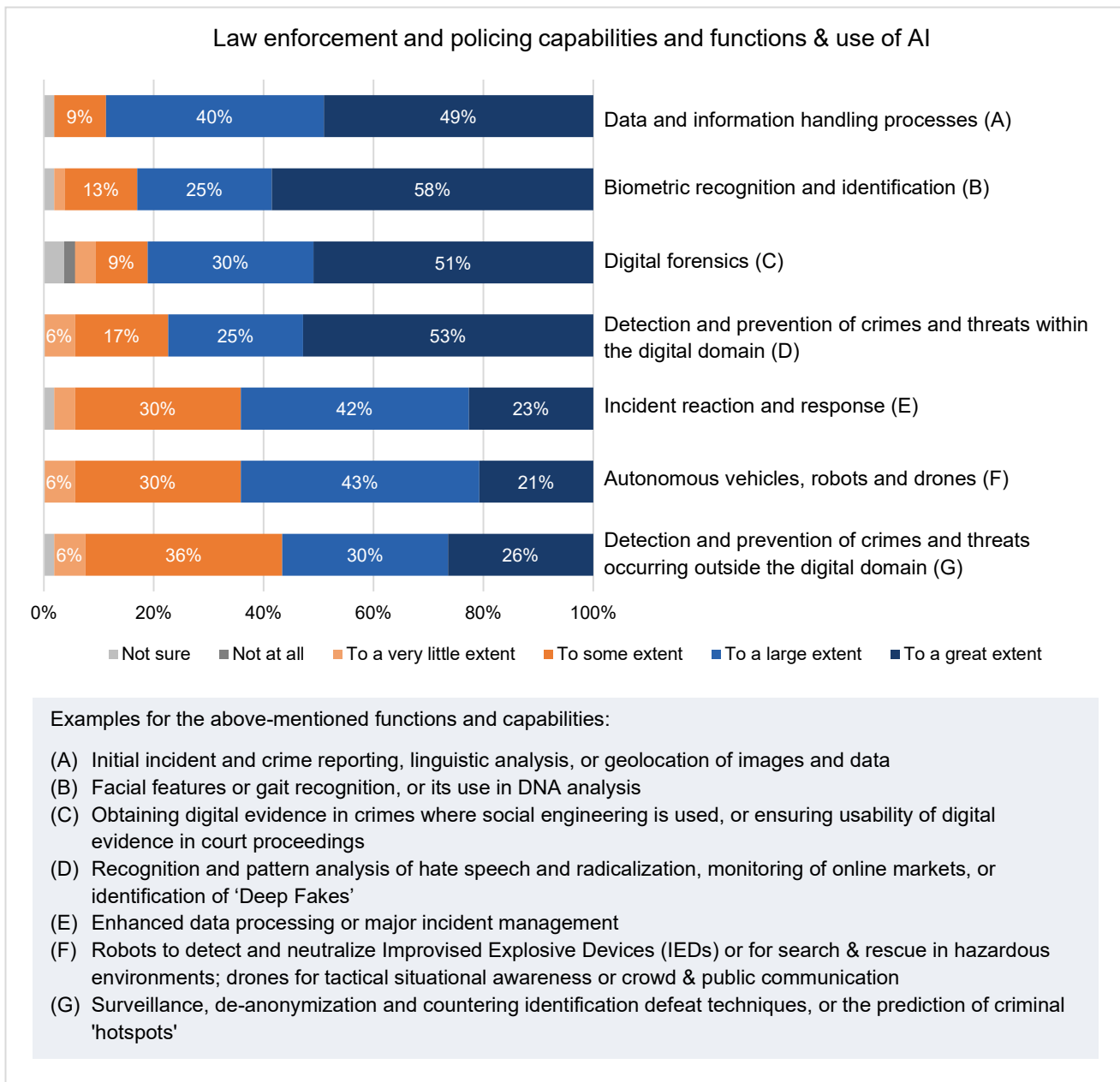


Figure 6: Results of the question “To what extent could the following law enforcement and policing functions and capabilities benefit from the use of AI?”, first ALIGNER survey, 2022.

However, potential does not necessarily imply immediate benefits. Therefore, ALIGNER also asked survey participants to identify work areas where AI could be used immediately to bring about consequential beneficial changes to P&LEA practice. Such an open question allowed the participants to formulate their views in their own words. The answers obtained were roughly clustered and prioritised, resulting in the following areas that were mentioned several (≥ 5) times:¹³ (i) Data and information handling processes; (ii) Digital forensics; (iii) Prevention of crimes within the digital domain, with a focus particularly on social media analysis; and (iv) Biometric identification. These are in line with the highest ranked work areas that exhibit the highest potential in general.

¹³ The full list of unclustered and unprioritized answers can be found in Annex B.



3.3 Perceived Challenges of AI in Law Enforcement and Policing

If AI has such a large potential for P&LEAs, why is it not already in broader use? What hinders the deployment of AI at P&LEA institutions? When asked these questions, workshop participants and respondents of the first survey¹⁴ brought up several challenges, which can be broadly categorized into

- ◆ **Ethical challenges** related to topics of discrimination, transparency, trust in the AI decision making process, and human oversight;
- ◆ **Legal challenges** related to safeguarding of fundamental rights, handling of AI system failures, privacy concerns, or ensuring usability of results from processes that make use of AI in court;
- ◆ **Institutional challenges** related to complicated procurement processes or difficulties in transferring promising outputs from research projects into practice; and
- ◆ **Technical challenges** related to the need to integrate new AI systems into legacy IT systems, the lack of appropriate training data, or the lack of knowledge and understanding of the underlying technology.

The ethical and legal dimensions of AI in law enforcement and policing were universally regarded as the most important issue, both by workshop and survey participants (see Figure 7 for an example quote). The most obvious issue here is certainly the compliance with fundamental rights, data protection, and privacy regulations, as AI systems usually require large amounts of data, which can easily result in (unintentional) mass surveillance. Other concerns relate to algorithm bias or the concern that AI is not used in a responsible way, e.g., fairly towards each citizen. In this context, the lack of trust in AI (presumably both among the public and among practitioners in P&LEAs) is mentioned several times as a challenge. Related to these concerns is the aspect that AI, when used by P&LEAs, must not replace the human brain or human decision making, e.g., in the interpretation of laws, as it is not considered capable of handling ‘the margin between right and wrong’.

“While AI can enhance capabilities [...], this does not mean it is a good use of AI for society.”

Figure 7: Quotation of one survey response which outlines the need for tradeoff between usefulness and the ethical and societal questions to explore.

In addition to ethical and legal concerns, one reason for the lack of trust could also be the lack of knowledge and understanding of the technology and thus the lack of transparency. Another technical challenge related to the lack of labelled training data for AI. This and the use of “bad quality” or “wrong” training data may then in turn have ethical and legal implications, such as creating algorithm bias.

Other important challenges mentioned are further legal issues, e.g., how to legally handle a failure of an AI system causing any kind of harm, and institutional issues, e.g., the degree of digitalisation of P&LEAs. In discussions with law enforcement practitioners, the complex procurement practices at public offices, the perceived aversion of top-level hierarchy towards AI systems, and general problems of transferring promising research results into practical use (e.g., because the technology developer does not provide support after the project ends) were also mentioned. These challenges are often compounded by a lack of resources (personnel, time, money) needed to develop, test, deploy, and run AI systems.

¹⁴ The full list of original responses can also be found in Annex B.



Perceived Challenges Related to the ALIGNER Narrative

Focusing on the AI systems and technologies highlighted in the ALIGNER narrative, various technical and organizational challenges became evident during workshops and discussions with practitioners.

AI systems for situational awareness – for example for incident response on the city-level – face issues such as data overload and insufficient training data. These systems may struggle to accurately interpret new situations, potentially resulting in errors. Biases and inaccuracies in AI assessments can lead to poor decision-making during critical events. Handling large datasets also raises privacy concerns and requires significant computational resources and specialized expertise, which are often limited. Additionally, the use of AI systems for situational awareness can result in ethical and legal quandaries. If the training data for these algorithms are biased or incomplete, the outcomes may unjustly target certain communities, raising questions about accountability and fairness, especially if wrongful actions are taken based on the AI outputs.

The complexity of veracity assessment – e.g. for authorship attribution to detect fake news – lies in analysing extensive data from multiple, unstructured sources and distinguishing between accurate information and misinformation. Legal challenges like privacy regulations and advanced encryption complicate deanonymization efforts. Authorship attribution is challenged by linguistic and behavioural variations, identity spoofing, and increased use of AI-generated text. Geolocating images is hindered by insufficient metadata and image alterations. Detecting deepfakes – necessary to counter disinformation and social manipulation – is becoming increasingly difficult due to their sophistication and the high processing costs involved.

To combat disinformation using deepfakes, P&LEAs require resources for deploying and training counter-AI systems. Geolocating images demands substantial amounts of labelled data. In addressing disinformation and social manipulation, practical issues exist in defining fake news and legally tackling it. Questions emerge such as when the distribution of fake news constitutes a crime, if ever. Some European countries have implemented legal and organizational tools to combat fake news (e.g., Germany, the Czech Republic, Hungary, and France), but experts and civil society representatives continue to express concerns about undermining free speech.

Addressing and using bot networks – often used in cybercrimes – introduces further ethical and legal complications, such as when the use or deployment of a bot network becomes a crime, if at all. There might be an ethical and legal justification for using bot networks. One approach could involve prebunking, also known as inoculation, to counter fake news. This method exposes individuals to a small amount of fake news to prepare them with arguments to defend against it, aiming to enhance resilience against external influences. However, this technique poses additional ethical and legal questions: who decides when to use prebunking techniques, and who supervises these procedures?



3.4 Perceived Trends in the Criminal Misuse of AI

In addition to the potential of AI for P&LEAs and the diverse challenges arising directly from the operational implementation of AI systems at P&LEAs, P&LEAs also need to address emerging trends in the criminal misuse of AI technology. Based on scenarios identified in the first two ALIGNER workshops, the project team asked practitioners from P&LEAs, research, industry, and policy about their perception regarding criminal use of AI technology during the first survey in 2022.

Results (see Figure 10) show that the largest criminal potential of AI is seen in disinformation and social manipulation (87% identify this as an area with large or great potential for criminal misuse), fraud and forgery (also 87%, but with fewer respondents identifying this as having great potential), data harvesting and exploitation (85%), exploitation of AI capabilities (75%), and social engineering (70%)¹⁵. The least potential is seen in weaponized/criminalized autonomous vehicles (41%) and AI-controlled civil and militarized robots (49%), likely due to the low penetration rate in day-to-day society.

Interestingly, the emergence of new AI applications has not yet resulted in a clear trend towards increased criminal use of AI (Figure 8), with nine (of 30) P&LEA respondents identifying some extent of an impact, 11 identifying no or not much impact, and 10 respondents having no clear picture. Respondents are also divided on whether criminals will gain the upper hand in utilizing AI against P&LEAs (Figure 9), with 14 respondents being rather in favour of this statement, 15 disagreeing, and one respondent being unsure.

Given this critical juncture, it is essential to establish suitable framework conditions and targeted research efforts to steer the expected arms race between criminals and P&LEAs in the right direction. This will help ensure that AI systems are used responsibly and ethically to enhance public safety and security.

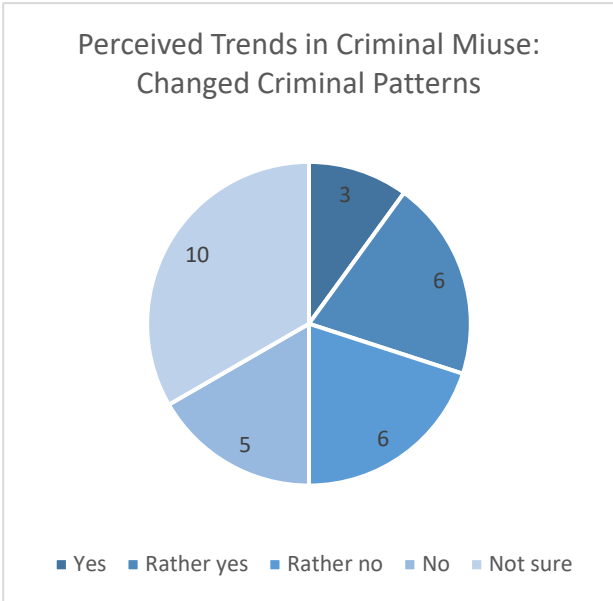


Figure 8: Results of the question "Have you noticed whether the emergence of new AI applications has an impact on criminal cases?", second ALIGNER survey, 2023

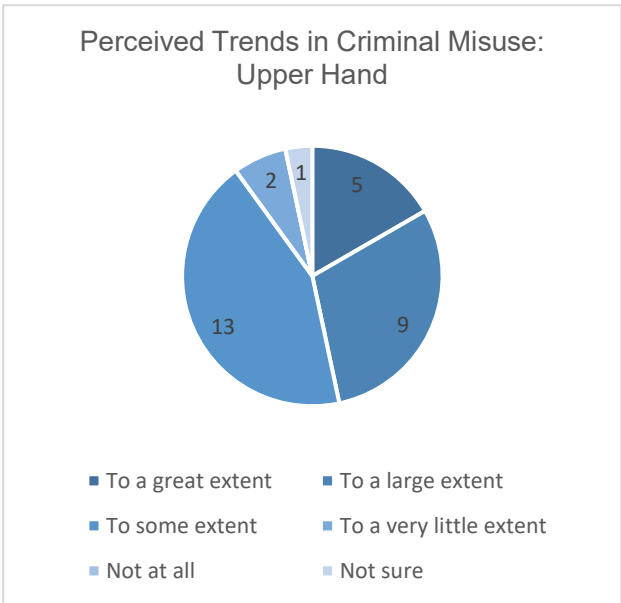


Figure 9: Results of the question "From a personal perspective, to what extent do you think criminals will have the upper hand in the use of AI against law and order?", second ALIGNER survey, 2023.

¹⁵ Results shown in Figure 10 show answers from all respondents.

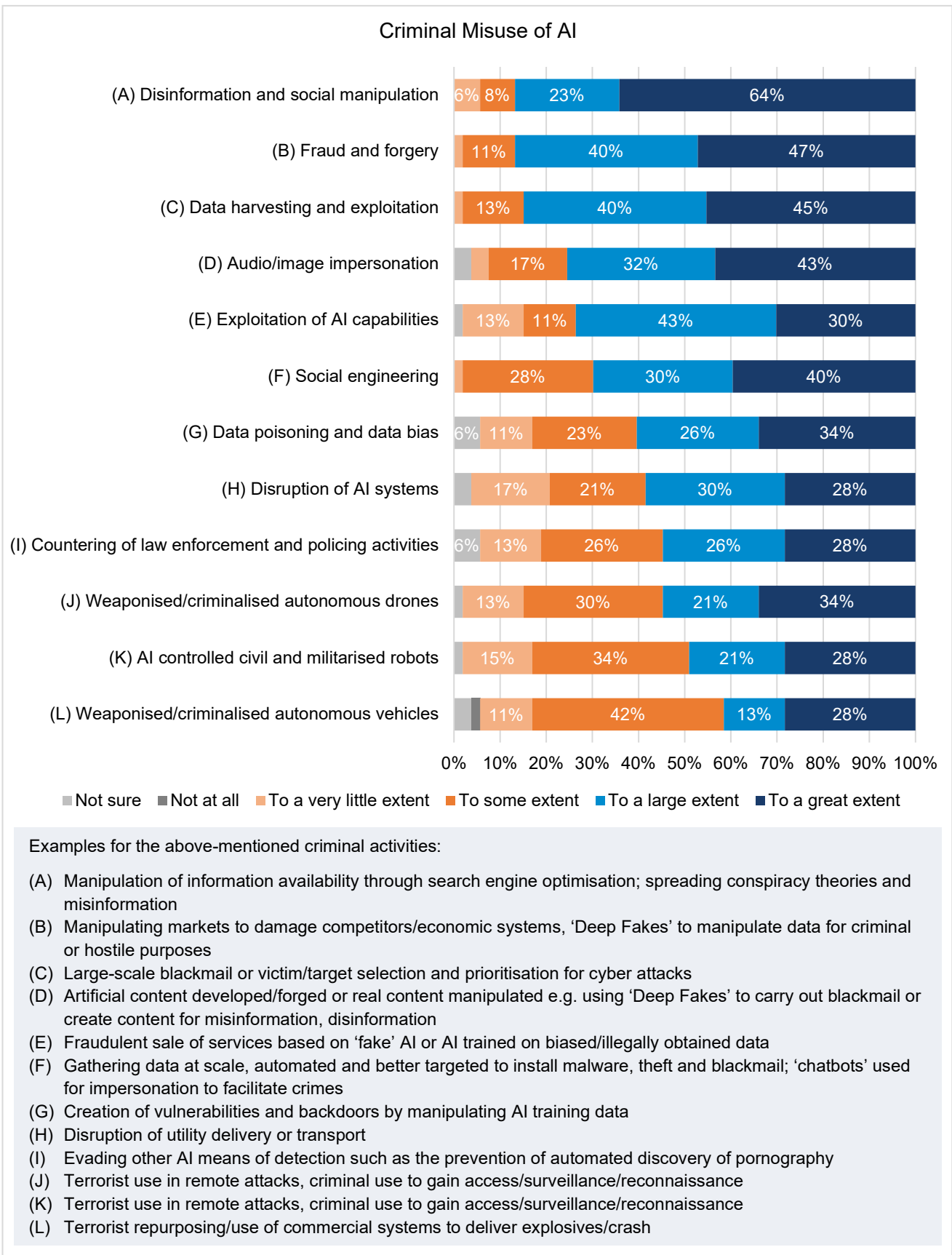


Figure 10: Results of the question "In your opinion, to what extent do the following criminal activities involving AI pose a security threat?", first ALIGNER survey, 2022.



4. Implications of Broader Distribution of AI Technology

The rapid expansion of AI technology across various sectors brings with it a plethora of implications. This section aims to delve into some of the multifaceted consequences of this widespread adoption that can currently be foreseen, focusing on three key areas:

- ◆ **Ethical and Legal Aspects:** The complex discussions surrounding the ethical deployment of AI by P&LEAs, which pose significant challenges to fundamental rights and legal frameworks.
- ◆ **Cyber Security Threats:** The risks associated with the integration of AI systems into our digital infrastructure, which can be vulnerable to cyber-attacks and data breaches.
- ◆ **Malicious Use of AI:** The potential for AI to be exploited for harmful purposes.

Together with the capability enhancement needs of P&LEAs, these points serve as the foundation for the policy and research recommendations that will be discussed in the succeeding sections.

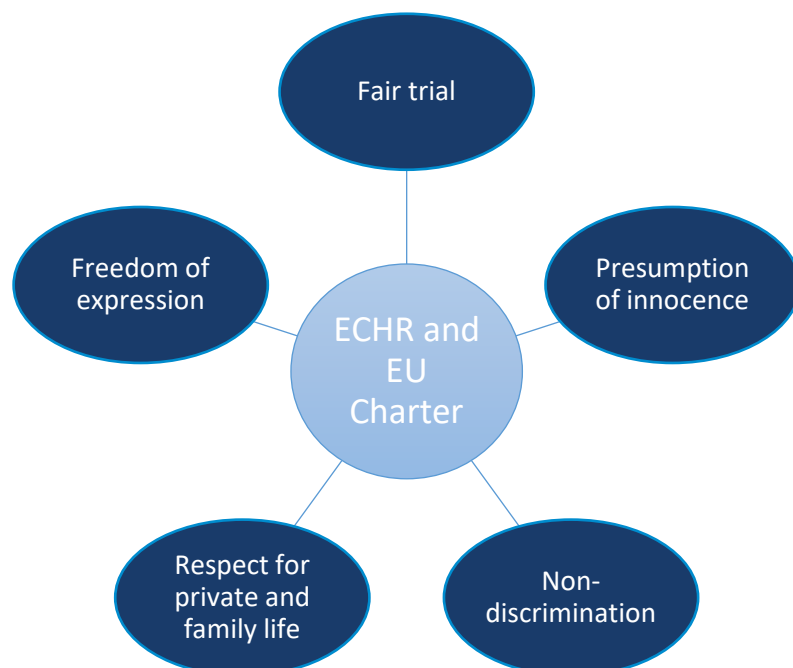
4.1 Ethical and Legal Aspects

While AI systems can bring clear benefits to the work of P&LEAs, they also raise numerous legal and ethical concerns, as already pointed towards in the previous section. If not properly developed and deployed by P&LEAs, these systems can significantly harm the fundamental rights of the concerned individuals. Therefore, it is crucial to specifically assess the potential risks that may arise from the P&LEAs' use of AI systems, and identify methods and best practices to prevent harm, well before the said systems are developed and deployed in practice.

Currently, many existing pieces of legislation focus on fundamental rights protection and establish obligations for state authorities that must be observed also by P&LEAs while deploying AI systems.

In the European Union (EU), fundamental rights of individuals are guaranteed and safeguarded by the two major human rights instruments adopted by the Council of Europe and the EU: The European Convention on Human Rights (ECHR) and the Charter of Fundamental Rights of the EU (the Charter). AI systems used for law

enforcement purposes are susceptible to affecting a multitude of fundamental rights guaranteed by the two instruments, as these rights are closely connected and intertwined. However, in the present context, the following fundamental rights are more likely to be impacted by the use of AI systems: the





presumption of innocence and the right to an effective remedy and a fair trial; the right to equality and non-discrimination; the right to respect for private and family life and the right to protection of personal data; and, finally, to freedom of expression and information.

For each of these rights, the relevant provisions of both the ECHR and the EU Charter as well as their further implications are summarized in the tables below. Additionally, the same tables show the potential harmful impact on fundamental rights of LEAs' use of AI systems, together with some suitable mitigation measures.

Presumption of innocence, right to an effective remedy and to a fair trial

Relevant provisions	Articles 6 and 13 ECHR and Articles 47 and 48 EU Charter.	
Definition & consequences	<p>Anyone charged with a criminal offence must be presumed innocent until proved guilty according to law.</p> <p>Anyone whose rights and freedoms are violated has the right to an effective remedy before a tribunal.</p> <p>Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law, including rights:</p> <ul style="list-style-type: none"> ◇ to be informed promptly of the nature and cause of the accusation; ◇ to bring their arguments and evidence as well as scrutinise and counteract the evidence presented against them; and ◇ to obtain an adequately reasoned and accessible decision. 	
AI-related risks	Mitigation measures	
<p>The AI system does not provide percentages or other indication on the degree of likelihood that the outcome is correct/incorrect, prejudicing the user that there is no possibility of error and therefore that the outcome is undoubtedly incriminating.</p>	<p>Assessment of the accuracy and reliability of the AI system deployed and communication of this information to the AI user.</p>	
<p>The AI system produces an outcome that forces a reversal of burden of proof upon the suspect, by presenting itself as an absolute truth, practically depriving the defence of any chance to counter it.</p>	<p>Ensuring meaningful human oversight and decision-making powers and that factual elements flagged by the AI system are not considered proven, unless supported by solid evidence.</p>	
<p>There is no explanation of reasons and criteria behind a certain output of the AI system that the user can understand.</p>	<p>Prosecution should be able to sufficiently explain the outputs generated by the AI systems used, to allow all relevant parties to challenge the evidence produced.</p>	
<p>The collection and preservation of AI-generated evidence is unlawful, leading to inadmissibility of the evidence in a criminal proceeding.</p>	<p>Ensuring lawful collection and preservation of chain of custody of AI evidence with appropriate safeguards.</p>	



Right to equality and non-discrimination

Relevant provisions	Article 14 ECHR and Articles 20 and 21 EU Charter.	
Definition & consequences	<p>Everyone is equal before the law. Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.</p> <ul style="list-style-type: none"> ◇ Everyone should be protected against discriminatory decisions or policies, including automated decision-making based on sensitive data. 	
AI-related risks	Mitigation measures	
The inaccuracy or non-exhaustivity of the criteria used in the design of the algorithm, as well as the poor quality or the existence of biases in the datasets used, leads the AI system to perpetuate or generate discriminatory outputs.	Testing and enhancing the quality and diversity of the datasets used to feed the AI system, to avoid biased outputs.	
	Avoiding the use of unlabelled datasets, to lower the risk that the new crime patterns or new criminal profiles identified are based on sensitive characteristics of the individuals.	
	Expanding the room for human intervention in both the design and deployment stages of the AI system, to minimise the risks of and correct inaccurate outputs.	

Right to respect for private and family life and right to protection of personal data

Relevant provisions	Article 8 ECHR and Articles 7 and 8 EU Charter.	
Definition & consequences	<p>Everyone has the right to respect for their private and family life, home and communications.</p> <ul style="list-style-type: none"> ◇ Self-development without state interference. <p>Everyone has the right to the protection of personal data concerning them.</p> <ul style="list-style-type: none"> ◇ Personal data must be processed fairly for specified purposes and on a legitimate basis. ◇ Rights of access and rectification. ◇ Independent oversight. 	
AI-related risks	Mitigation measures	
The use of large datasets including a vast amount of personal and sensitive data causes a	Where possible, opting for synthetic datasets or anonymised datasets with lowest risks of re-identification.	



disproportionate interference with privacy and data protection rights.	Performing a data protection impact assessment, to assess the legality and proportionality of the interference and strict adherence to data protection principles and relevant secondary legislation.
Continuously merging and repurposing different datasets leads to the development of mass surveillance tools and chilling effects.	Avoiding the repurposing of datasets and limiting their use to the original purpose foreseen during the data collection.
There are no governance procedures to ensure the lawfulness of the processing of personal data, limit the amount of personal data processed and comply with data subject rights.	Designing the personal data processing operations by adopting a data protection by design and default approach to ensure data protection compliance throughout the entire processing lifecycle.

Freedom of expression and information

Relevant provisions	Article 10 ECHR and Article 11 EU Charter.
Definition & consequences	<p>Everyone has the right to freedom of expression, including freedom to hold opinions, communicate and acquire information</p> <ul style="list-style-type: none"> ◇ State negative obligation not to interfere and positive obligation to facilitate the exercise of the right
AI-related risks	Mitigation measures
The deployment of the AI system is not limited to suspected individuals.	Implementing mechanisms to limit the deployment of the AI systems to (groups of) individuals on the basis of suspicions or objective criteria.
AI-enabled surveillance systems lead to a chilling effect especially for minority groups, who refrain from expressing their opinions.	Avoiding a targeted use of such tools on minorities and marginalised communities, to not deter them from publicly expressing their opinions.
Data stored and recorded by the AI system are non-easily accessible for individuals who want to exercise their right to information.	Ensuring the information stored by the AI systems is available, understandable and easily exportable.



4.2 Cybersecurity Aspects of AI

The extensive use of AI in everyday tasks increases the chances of malware being hidden within AI systems or their data sets. Some notable cybersecurity threats posed by AI systems include injection of malicious code, reconstruction of training data, instruction extraction, communication extraction, knowledge base extraction, model performance degradation, indirect prompt injections, and the compromise of data brokers/providers. Below, we present key examples of possible cybersecurity threats posed by AI systems, as detailed during ALIGNER workshops and emphasized by the German Federal Office for Information Security (BSI) [2, 3] and the European Union Agency for Cybersecurity (ENISA) [4]. For more details on cybersecurity threat assessment frameworks, we refer to [5].

Injection of malicious code is a prevalent threat as LLMs are increasingly used to generate program code or refer to third-party sources. This allows attackers to strategically place their malicious code in public program libraries, aiming for these libraries to be recommended by LLMs. When executed, this code can exfiltrate sensitive information, impair system availability, or break out of a sandbox environment.

Reconstruction of training data is another significant threat. Attackers can reconstruct a model's training data through targeted queries, even if the data appears infrequently. Membership inference attacks can determine whether specific data or documents were part of the LLM's training material, which is particularly critical if the training data was extracted from the internet without thorough examination.

Instruction extraction involves attackers attempting to extract instructions that precede user inputs to the LLM. These instructions guide the model's behavior and can be exploited by attackers for **prompt injections**. **Communication extraction** targets chat histories between users and chatbots, aiming to extract sensitive information.

Knowledge base extraction seeks to extract information stored in knowledge bases accessible to LLMs. This includes documents used to substantiate the LLM's outputs, posing a risk if sensitive or proprietary information is included.

Model performance degradation or model poisoning involves attackers deceiving the LLM with minor input changes, degrading its performance. For instance, an attacker can introduce a pixel pattern in training data labeled as "policeman," allowing them to bypass surveillance systems by wearing a t-shirt with the pattern.

Indirect prompt injections exploit third-party sources to manipulate LLM behavior. Attackers hide instructions on websites, in emails, or documents evaluated by the LLM, influencing the conversation or triggering intensive queries that can slow down the system.

The **compromise of data brokers/providers** is also a significant threat. Attackers can manipulate the data sent to the AI process, poisoning datasets or deleting registries, impacting the accuracy and reliability of learning models.

To counter these cybersecurity threats to AI systems, P&LEAs should consider the countermeasures proposed by the BSI [3] and forthcoming guidelines from the AI Office. These recommendations provide a strong foundation for developing strategies to mitigate the risks associated with AI systems and ensure their secure and ethical deployment.



4.3 Potential Malicious Use of AI

Based on the results of the first two ALIGNER workshops and surveys, as well as discussions during further workshops between 2022 and 2024, several potential examples of the malicious use of AI for criminal activities were identified. These examples were clustered around three different threat categories, corresponding – as far as possible – with the topics of the ALIGNER narrative: Disinformation and social manipulation, cybercrime against individuals and organizations, and vehicles, robots, and drones. The following sections provide an overview of these examples to provide P&LEA practitioners, researchers, and policymakers with a starting point to develop counter strategies for the malicious use of AI. For more detail on the development of this taxonomy, please see [6].

4.3.1 Disinformation and Social Manipulation

Threat	Potential crime	Example for malicious AI use
Data extraction	Extortion	AI may be used to systematically harvest data about companies, individuals and the government for tracking, manipulation and extortion.
	Improper activity at election	Deep fakes may be used to misinform the public. For example, it may be used to influence politics and elections (i.e. by releasing fake audio or video recordings of political figures).
AI deep fakes	Nonconsensual pornography' and child pornography	Criminal use of deep fakes as child pornography and/or non-consensual pornography.
	Extortion	Criminals and/or terrorist groups may use deep fakes to trick, threaten and extort people to raise funding. Criminals can also use deep fakes to trick people in critical positions to collect and reveal classified, confidential or personal information.
	Incitement of violence	Deep fakes can be used to fabricate politicians that incite people to act in a harmful way. This could for example lead to situations like the US Capitol riots in January 2021.
	Information theft	Criminals and terrorists may use deep fakes to impersonate people in critical positions to obtain critical and perhaps confidential or classified information.
	Reputational damage	Criminals may use deep fakes to impersonate people to damage their reputation, e.g. by falsely attributing controversial opinions or false information.



Biometric spoofing	Identity theft	Biometric uses attributes such as voice, fingerprints and handwriting to identify individuals. Today, verification of people having access to phones, buildings etc. is possible with the use of biometrics. Criminals may create new biometrics samples to hack systems or to generate spoof handwriting or synthetic fingerprints.
Fake evidence	Extortion	AI may be used to automatically collect evidence or produce fake evidence to up-scale extortion.
Influence campaigns	Incitement of violence	AI can contribute to an increased spread of terrorist or violent narratives that can incite people to act in a harmful way.
Information campaigns	Incitement of hate speech	AI can make information operations more scalable, precise and persistent. Malign information is already an existing problem but can be aggravated with use of AI. For example, AI can be used to manipulate content or produce content to manipulate messages and spread malign information by embedding AI into different platforms. The information can be used to incite people to act in a harmful way.
Denial-of-information attacks	Fraud & Forgery	AI supported bot-driven, large-scale information-generation attacks can be used to making it more difficult to obtain correct information. The attacks may be used to target military, economic and educational infrastructure to make correct and vital information harder to access.
Social engineering attack	Fraud & Forgery	A victim's online information can be used to automatically generate custom malicious websites/emails/links the victim would be likely to click on, so called spear phishing. The communication is sent from addresses that impersonate their real contacts, using a writing style that mimics those contacts.
	Swindling	Phishing attacks can be improved by using AI to construct messages that appear more genuine. AI techniques can be used for active learning to discover the work that will result in maximized responses by varying the details of messages to gather data. The scalability and frequency of an attack can be improved by e.g. spear phishing where AI can create more effective and extensive attacks.



Fake news	Incitement of violence	Fake news reports can be used to fabricate persons that a victim trust. For example, fake news can incite people to act in a harmful way.
Hacking	Breach of data secrecy	The computerization of diverse fields, from finance to elections, increases the speed, scale, and scope of vulnerability to hacking. AI can be used to evade detection, improve target selection, improve prioritization, and creatively respond to changes in the target's behavior. For example, AI can be used to destruct and disclose personal data.

4.3.2 Cybercrime Against Individuals and Organizations

Threat	Potential crime	Example for malicious AI use
Adversarial AI	Fraud & Forgery	Cybercriminals can use AI techniques to automate various tasks, such as dialogue with ransomware victims, payment processing and facilitate medical insurance fraud. For example, research has displayed how adversarial attacks in the healthcare sector can be carried out using co-opt diagnostic algorithms.
	Information theft/ Espionage	Cybercriminals can use AI techniques to steal information and expose it. By using “exploratory attacks” criminals can extract information (for example training data) from AI models.
	System interference	Attacks against machine learning can be used to commit system interference, e.g. “evasion attacks” which are conducted by creating malicious inputs that may generate a false prediction for the model.
	Breach of data secrecy	Criminals can use poisoning attacks (that aim to create backdoors in consumer ML and/or generate surreptitiously harm) to commit crimes such as data interference. Even small manipulations of algorithms or data sets can lead to substantial changes for how AI systems operate.
Denial of services (DDoS)	Breach of data secrecy	AI supported DDoS attacks may be used to target military, economic and educational infrastructure to withhold information.
Malware	Information theft	Criminals can use AI to create malware (malicious software), for example to obtain confidential information.



	Extortion	Criminals can use AI to create ransomware (a type of malware) to extort money from victims but can also be used with destructive or disruptive purposes as seen in the NotPetya attack in 2017 ¹⁶ .
	Sabotage	Criminals can use AI to create malware worms to sabotage infrastructure and operative systems. This was done in the case of Stuxnet in 2010.
Fake news	Improper activity at election	Fake news reports with realistic fabricated audio and video of state leaders can be interpreted as realistic causing people to act or vote differently than otherwise. For example, deep fakes of candidates for elections may impact the outcome of the voting where the technique can be used to undermine confidence in the individual politician or party they represent.
	Incitement of violence	See example in previous category
Social engineering attack	Swindling	Phishing attacks can be improved by using AI to construct messages that appear more genuine. AI techniques can be used for active learning to discover the work that will result in maximized responses by varying the details of messages to gather data. The scalability and frequency of an attack can be improved by e.g. spear phishing where AI can create more effective and extensive attacks.
	Fraud & Forgery	A victim's online information is used to automatically generate custom malicious websites/emails/links the victim would be likely to click on, so called spear phishing. The communication is sent from addresses that impersonate their real contacts, using a writing style that mimics those contacts.
Password guessing	Information theft	AI can be used to expedite, enhance and automate the process of password guessing. By obtaining passwords and access protected websites, malicious actors can enter systems or networks, to create disruption, disrupt essential services, steal information and/or data, manipulate data or processes or install malicious software.
CAPTCHA breaking	Breach of data secrecy	CAPTCHA is a security measure used to protect networks and websites from various attacks. Criminals can carry out cyberattacks by using AI to overcome CAPTCHA.

¹⁶ <https://www.cloudflare.com/learning/security/ransomware/petya-notpetya-ransomware/> [accessed 2024-09-22]



Market bombing	Swindling	AI can be used to manipulate financial or stock markets via targeted, high frequency, patterns of trades, to harm currencies, competitors or the economic system. A side effect can also be that AI creates profit from trading even if that is not the direct yield.
AI supported crypto currency trading	Fraud & Forgery	AI can manipulate cryptocurrency for financial profit.
	Theft	Criminals could facilitate theft of cryptocurrencies by using AI techniques.
Tricking face recognition	Identity fraud	AI systems are used for face recognition which could be used as ways of tricking identification systems resulting in identity fraud.
Online stalking	Persecution	AI can improve discovering and monitoring individuals' activities and through personal device data or social media. This increases the possibility to stalk the targeted individuals.
Automated surveillance platforms to suppress dissent	Violation of human rights	States may use automated audio and image processing to extend state surveillance in an unproportionate way or to suppress debate.

4.3.3 Vehicles, Robots, and Drones

Threat	Potential crime	Example for malicious AI use
Weaponized or criminalized autonomous vehicles	Terrorism	Commercial systems can be used in harmful and unintended ways, e.g. autonomous vehicles may deliver explosives and cause crashes.
	Traffic violation	
	Harmful explosion/Arson	
	Drug trafficking	Criminals may use autonomous vehicles for drug trafficking. For example, US law enforcement actors have discovered and seized autonomous underwater vehicles used for drug trafficking.
AI-controlled robots for	Physical assault	AI-controlled robots can be used to carry out a physical attack. This may unlawfully cause, e.g., injury, damage or destruction.



harmful or malicious use	Harmful explosion/ Arson	Criminals may use military robotics research and its inventions to commit crimes. For example, robotics could be used to deliver explosives causing a harmful explosion/arson.
Weaponized or criminalized autonomous drones	Harmful explosion/arson Drug trafficking and/or drug dealing	Drones can be used in several harmful ways, whether originally designed for it or not. For example, drones can deliver explosives. Criminals can use drones for drug trafficking or drug dealing. They may also be used to facilitate smuggling.



5. Policy and Research Recommendations

Building on the opportunities, challenges, and risks associated with the (mis)use of AI, ALIGNER has developed comprehensive recommendations for policymakers, decision-makers, researchers, and the industry. These recommendations are aimed at addressing the operational, cooperative, and collaborative needs of P&LEAs. The recommendations have been formulated in the context of ongoing policy processes, particularly the AI Act, which was drafted, debated, and enacted by the European Commission during the project's duration. This section first provides essential information on the relevant legal framework, setting the stage for the subsequent policy and research recommendations.

5.1 The Relevant Legal Framework: The AI Act

On 21 April 2021, the European Commission published its proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence, commonly known as the 'AI Act' [7]. After a highly debated legislative procedure that lasted more than three years, the AI Act became law on 13 June 2024 and entered into force on 1 August 2024 [8].

Pursuant to its Article 1, the AI Act has two main objectives: (1) improving the functioning of the internal market and promoting the uptake of human-centric and trustworthy AI by laying down a uniform legal framework for the development, placing on the market, putting into service, and use of AI systems in the EU; and (2) ensuring a high-level of protection for health, safety, and fundamental rights enshrined in the Charter. As such, the AI Act combines both product safety regulation and fundamental rights protection [9], as shown by its two legal bases: (1) Article 114 of the Treaty on the Functioning of the EU (TFEU), which provides the EU with the power to adopt measures for the approximation of legal provisions on the establishment and functioning of the internal market; and (2) Article 16 of the TFEU, which provides the EU with the power to adopt rules on the protection of individuals with regard to the processing of their personal data.

The AI Act is a regulation, namely a legal act of general application, binding in its entirety and directly applicable to all EU Member States [10]. Thus, the AI Act applies as it is and no national transposition is required or allowed, unless when explicitly stated otherwise [11]. Moreover, the AI Act pursues a horizontal regulatory approach, by setting a comprehensive but flexible enough legal framework to be future-proof and apply to all AI systems in the EU and throughout their entire lifecycle [7].

The following sections illustrate the AI Act's scope of application (5.1.1) and its risk-based approach (5.1.2).

5.1.1 Scope of application

The AI Act's horizontal regulatory approach is reflected in the broad and flexible definition of 'AI system', included in its Article 3: an AI system is *"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"*. According to Recital 12, the definition puts an accent to those key characteristics distinguishing an AI system from a simpler traditional software, such as the capability to operate according to explicit or implicit objectives or its autonomy and adaptiveness. In particular, Recital 12 clarifies that of utmost importance is the AI



system's capability to infer, namely the process of obtaining outputs, models, and/or algorithms from inputs or data, which *“transcends basic data processing by enabling learning, reasoning or modelling”*.

According to its Article 2, the AI Act's personal scope of application covers two main categories of entities: providers and deployers of AI systems. According to Article 3(3) of the AI Act, a provider is *“a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge”*. Thus, a provider is the entity responsible for placing on the market or putting into service an AI system directly developed or commissioned by them. As a result, the AI Act does not apply to any research, testing (unless happening in real world conditions), or development activities prior to the AI system being placed on the market or put into service. According to Article 3(4) of the AI Act, a deployer is *“a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity”*. Thus, a deployer is the entity responsible for using during a professional activity an AI system developed by others.

According to its Article 2, the AI Act's geographical scope of application broadly extends to both AI systems providers and deployers established or located within the EU and providers and deployers established or located in a third country, when the AI system is put into service in the EU or placed on the EU market or when the output produced by the AI system is used in the EU.

Article 2 also introduces specific exceptions to the AI Act's scope of application. The AI Act does not apply to: (1) areas outside of the scope of EU law, such as national security, which remains an exclusive competence of Member States; (2) AI systems placed on the market, put into service, or used exclusively for military, defence or national security purposes; (3) public authorities of third countries and international organisations where the use is carried out in the framework of agreements for law enforcement or judicial cooperations with the EU or one or more Member States, as long as adequate safeguards to the protection of fundamental rights are established; and (4) AI systems developed and put into service for the sole purpose of scientific research and development.

5.1.2 Risk-based approach

To avoid creating unnecessary restrictions and burdens to providers and deployers of AI systems, the AI Act adopts a risk-based regulatory approach limiting the legal intervention to those situations where there are justified concerns needing to be addressed [7]. In line with this, the nature and content of the obligations imposed on providers and deployers is proportional to the risks to

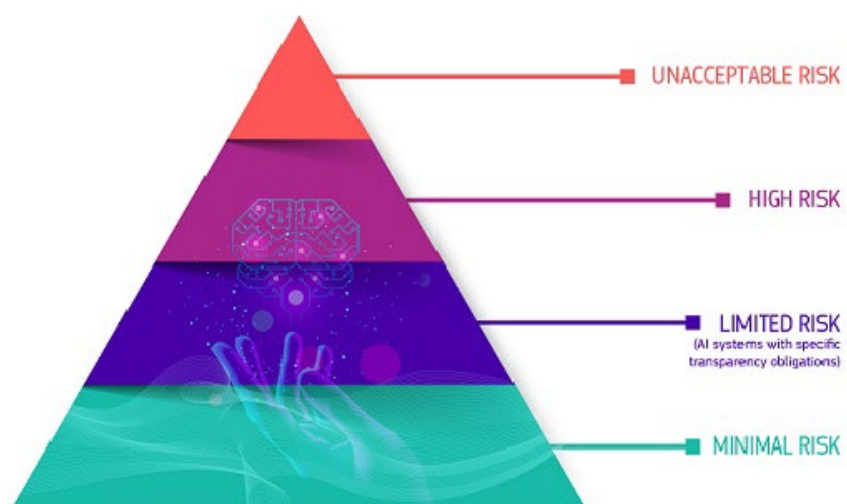


Figure 11: AI Act risk-based approach. Source: [24]



health and safety, as well as to fundamental rights, created by different categories of AI systems in different circumstances of use.

As shown in Figure 11, the AI Act distinguishes between four broad risk levels: (1) AI systems creating unacceptable risks, the use of which is prohibited; (2) AI systems creating high risk, subjected to mandatory requirements for providers and deployers and to an *ex-ante* conformity assessment; (3) AI systems creating limited risks, which are subjected to minimal transparency obligations; and (4) AI systems creating minimal risks, which can be freely placed on the market, put into service and used.

5.1.2.1 Prohibited AI systems

Article 5 of the AI Act identifies eight AI practices creating unacceptable risks, as their use is deemed incompatible with EU values and fundamental rights [7]. Thus, as of 2 February 2025, the placing on the market, putting into service, or use of AI systems falling within these practices will be prohibited.

Of particular relevance in the law enforcement context are the fourth, seventh, and eight prohibitions. The fourth prohibition relates to AI systems for “***making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics***”. As clarified by Recital 42, pursuant to the presumption of innocence enshrined in Article 48 of the Charter of Fundamental Rights of the EU, natural persons should never be deemed involved in criminal activities only on the basis of their (AI-predicted or -inferred) profiles, personality traits or characteristics such as nationality, place of birth, place of residence, economic situation. Thus, such predictive policing AI systems evaluating the likelihood of committing or predicting the occurrence of criminal offences are prohibited. However, these AI systems can still be used by LEAs and the judiciary to support human assessments of the involvement of a person in a criminal activity, after reasonable suspicions already based on objective and verifiable facts directly linked to a criminal activity.

The seventh prohibition relates to “***biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation***”. For the purposes of the AI Act, Article 3(34) includes a broad definition of ‘biometric data’ as personal data resulting from specific technical processing relating to physical, physiological, or behavioural characteristics of a natural person. When biometric data is used to infer the special categories of personal data listed, the AI system is prohibited unless used to label or filter lawfully acquired biometric datasets, also in the area of law enforcement.

The eight and last prohibition relates to **biometric identification systems**. To be prohibited, these AI systems need to: (1) operate in real-time, meaning that the capturing of the biometric data, the comparison with a reference database, and the identification occur without a significant delay (Article 3(42) of the AI Act); (2) be deployed in a publicly accessible space, meaning any publicly or privately owned physical place accessible to an undetermined number of natural persons, regardless of whether certain conditions for access may apply (Article 3(44) of the AI Act); and (3) be deployed for the purposes of law enforcement, namely the prevention, investigation, detection, or prosecution of criminal offenses or the execution of criminal penalties, including safeguarding against and preventing threats to public security (Article 3(46) of the AI Act).

As stated in Recital 32, these AI systems are particularly intrusive to and may have chilling effects on the fundamental rights and freedoms of individuals, as they may affect the private life of a large part of



the population by evoking a feeling of constant surveillance. Moreover, in case of biased results, these AI systems may lead to serious discriminatory effects, exacerbated by the immediacy of the impact and the limited opportunities for further checks or corrections by human operators. For these reasons, the use of these AI systems by LEAs is allowed only when strictly necessary in the following narrowly defined situations, all characterised by the existence of a substantial public interest: (1) targeted search for specific victims of abduction, human trafficking or sexual exploitation, as well as for missing persons; (2) prevention of a specific, substantial, and imminent threat to the life of physical safety of natural persons or a genuine and present or foreseeable threat of a terrorist attack; and (3) localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for the offences listed in Annex II to the AI Act¹⁷ and punishable in the Member State by a custodial sentence or a detention order for a maximum period of at least four years.

Real-time remote biometric identification systems deployed for any of these objectives can only be used by LEAs to confirm the identity of the specifically targeted individual. Before deciding to initiate the use, LEAs need to evaluate the nature of the situation giving rise to the possible use, specifically its seriousness, probability, and scale of harm in case the system were not used, as well as the consequences, their seriousness, probability, and scale caused by the use of the system for the fundamental rights of the concerned individuals. Prior initiating the use, LEAs need to request and obtain an *ad hoc* authorisation from the competent judicial or administrative authority. However, in duly justified situations of urgency, the use of the AI system may be commenced without such authorisation, which needs to be requested within 24 hours; in case of rejection, LEAs need to stop using the AI system and delete the data and outputs. The use can be authorised only where LEAs have already conducted a fundamental rights impact assessment of the AI system. The use by LEAs needs to comply with necessary and proportionate safeguards and conditions concerning, for instance, temporal, geographic, and personal limitations.

5.1.2.2 High-risk AI systems

Article 6 of the AI Act identifies two categories of AI systems creating high risks, as their functions, purposes, and modalities of use may adversely affect the health, safety, and fundamental rights of individuals [7]. Thus, the placing on the market, putting into service or use of high-risk AI systems will be subjected to compliance with mandatory requirements and an *ex-ante* conformity assessment.

Of particular relevance in the law enforcement domain, the second category of high-risk AI systems includes stand-alone AI systems with fundamental rights implications listed in Annex III to the AI Act. Annex III lists specific use-cases falling in the following eight areas: biometrics; critical infrastructure; education and vocational training; employment, workers' management and access to self-employment; access to and enjoyment of essential services and benefits; law enforcement; migration, asylum and border control management; and administration of justice and democratic processes. According to Article 7 of the AI Act, the European Commission can adopt delegated acts to amend Annex III by

¹⁷ Annex II lists the following offences: terrorism; trafficking in human beings; sexual exploitation of children, and child pornography; illicit trafficking in narcotic drugs or psychotropic substances; illicit trafficking in weapons, munitions or explosives; murder; grievous bodily injury; illicit trade in human organs or tissue; illicit trafficking in nuclear or radioactive materials; kidnapping, illegal restraint or hostage-taking; crimes within the jurisdiction of the International Criminal Court; unlawful seizure of aircraft or ships; rape; environmental crime; organized or armed robbery; sabotage; and participation in a criminal organization involved in one or more of the offences listed above.



adding or modifying use-cases, where the use-cases fall within the listed eight areas, and they pose an equivalent or greater risk than the risks posed by the already listed AI systems.

The stand-alone AI systems listed in Annex III to the AI Act will be subjected to the relevant obligations for high-risk AI systems as of 2 August 2026. However, according to Article 6(3) and (4) of the AI Act, a provider which considers that an AI system listed in Annex III does not pose a significant risk of harm can be exempted from the high-risk obligations by documenting this assessment and, upon request, providing it to the national competent authorities. The provider's assessment needs to focus on whether the AI system does not materially influence the outcome of decision-making; this particularly refers to the following situations: (1) the AI system performs a narrow procedural task¹⁸; (2) the AI system improves the results of a previously completed human activity¹⁹; (3) the AI system detects decision-making patterns or deviations from them²⁰; and (4) the AI system performs a preparatory task to an assessment²¹

Most relevant for LEAs are the first, sixth, and seventh areas and related use-cases listed in Annex III to the AI Act. The first area relates to **biometrics**, in so far as the use of AI systems for biometric purposes is permitted. Annex III lists the following use-cases: (1) remote biometric identification systems, as technical inaccuracies may lead to biased results and discriminatory effects; (2) biometric categorisation according to (inferred) sensitive or protected attributes, pursuant to Article 10 of the Law Enforcement Directive; and (3) AI systems used for emotion recognition, due to their questionable scientific basis and lack of reliability.

The sixth area relates to **law enforcement**, as the role, responsibility, and actions of LEAs involving AI systems are characterised by a significant degree of power imbalance and may lead to surveillance or deprivation of liberty, which may be unjust, discriminatory, and difficult to challenge and remedy, as reminded by Recital 59 of the AI Act. Annex III lists the following use-cases: (1) AI systems used to assess the risk of a natural person becoming the victim of a criminal offence; (2) polygraphs or similar tools; (3) AI systems used to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences; (4) AI systems used for assessing the risks of a natural person (re-)offending not solely based on profiling, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups; (5) AI systems used to profile natural persons in the course of the detection, investigation, or prosecution of criminal offences. All these AI systems are considered high-risk when used by or on behalf of LEAs and by EU institutions, bodies, offices, or agencies (e.g. Europol).

The seventh area relates to **migration, asylum and border control management**, as the affected individuals are often in a particularly vulnerable position and dependent on the outcome of the actions of the competent public authorities and the use of AI systems may lead to non-transparent results undermining the rights to free movement and non-discrimination, as reminded by Recital 60 of the AI Act. Annex III lists the following use-cases: (1) polygraphs or similar tools; (2) AI systems used to assess a risk, including a security risk, risk of irregular migration, or a health risk posed by a natural person who intends to enter or has entered into the territory of a Member State; (3) AI systems used to for the examination of applications for asylum, visa, or residence permits and associated complaints; (4) AI systems used for the purpose of detecting, recognising, or identifying natural persons. All these AI

¹⁸ For instance, transforming unstructured data into structured data.

¹⁹ For instance, improving the language of a previously drafted document.

²⁰ For instance, checking *ex post* whether there is a deviation in grading patterns.

²¹ For instance, file handling.



systems are considered high-risk when used by or on behalf of competent public authorities and by EU institutions, bodies, offices, or agencies (e.g., Frontex).

Pursuant to Article 26 of the AI Act, deployers of high-risk AI systems need to implement the appropriate technical and organisational measures to ensure that their use of high-risk AI systems complies with the instructions for use received by the provider. Where deployers consider that the use of the high-risk AI system in accordance with the instructions may result in a risk to health, safety, or fundamental rights, or where a serious incident happens, they need to immediately inform the provider. Deployers need to assign human oversight to individuals who have the necessary competence, training, authority, and support. If exercising control over input data, deployers need to ensure that input data is relevant and sufficiently representative in view of the intended purpose of the high-risk AI system. Deployers need to keep the logs automatically generated by the AI system for an appropriate period. Deployers of high-risk AI systems listed in Annex III to the AI Act need to inform individuals that they are subjected to a decision taken by automated means.

Pursuant to Article 26(10) of the AI Act, LEAs deploying post-remote biometric identification systems need to request and obtain, no later than 48 hours after commencing the use, an *ad hoc* authorisation from the competent judicial or administrative authority, unless when the system is used for the initial identification of a potential suspect based on objective and verifiable facts directly linked to the offence. In case of rejection, LEAs need to stop using the AI system and delete all the data and outputs. The use of the system needs to be limited to what strictly necessary for the investigation of a specific (threat of) criminal offence or for the search for a specific missing person.

Pursuant to Article 27 of the AI Act, deployers of AI systems listed in Annex III to the AI Act that are bodies governed by public law or private entities providing public services, thus including LEAs, need to perform a fundamental rights impact assessment of their high-risk AI systems, prior to commencing the use. The fundamental rights impact assessment consists of: (1) a description of the processes in which the AI system will be used; (2) a description of the period of time and frequency with which the AI system will be used; (3) the categories of individuals and groups likely to be affected by the use of the AI system; (4) the specific risks of harm likely to have an impact on the subjected individuals; (5) a description of the implementation of human oversight measures; and (6) the measures to be taken in case of materialisation of risks. To facilitate compliance with this obligation, the AI Office is expected to develop a template for a questionnaire. Meanwhile, ALIGNER developed the ALIGNER Fundamental Rights Impact Assessment templates, a tool specifically addressed to LEAs aiming to deploying high-risk AI systems for law enforcement purposes within the EU [12].

5.1.2.3 Limited- and minimal-risk AI systems

Article 50 of the AI Act identifies three categories of AI systems creating limited risks, as they may pose specific risks of manipulation, impersonation, or deception, as stated by Recital 132. Thus, as of 2 August 2026, these AI systems will be subjected to minimal transparency obligations.

The first category of limited-risk AI systems includes systems intended to **interact directly with natural persons**. Pursuant to Article 50(1) of the AI Act, providers of such AI systems need to design and develop them in such a way that the individuals concerned are informed that they are interacting with an AI system, unless obvious. This obligation does not apply to AI systems authorised by law to detect, prevent, investigate, or prosecute criminal offences.



The second category of limited-risk AI systems includes systems **generating synthetic audio, image, video, or text content**. Pursuant to Article 50(2) of the AI Act, providers of such AI systems need to design and develop them ensuring that their output is marked in a machine-readable format and detectable as artificially generated or manipulated. Pursuant to Article 50(4) of the AI Act, deployers of such AI systems need to disclose that the content has been artificially generated or manipulated. This obligation does not apply to AI systems authorised by law to detect, prevent, investigate, or prosecute criminal offences.

The third category of limited-risk AI systems includes **emotion recognition and biometric categorisation systems**. Pursuant to Article 50(3) of the AI Act, deployers of such AI systems need to inform the exposed individuals of the operation of the systems. This obligation does not apply to AI systems permitted by law to detect, prevent, investigate, or prosecute criminal offences.

All the AI systems not explicitly regulated by the AI Act are considered posing minimal risks to the health, safety, and fundamental rights of individuals. Thus, these AI systems can be developed and used subject to the existing legislation, without any additional legal obligations established by the AI Act [13].

5.2 Policy Recommendations

Based on those ongoing policy processes, discussions with experts from P&LEAs, research (including ethicists), industry, and policy during ALIGNER workshops between 2021 and 2024, as well as results from research and policy events jointly conducted with the EU AI cluster (ALIGNER, popAI, STARLIGHT, AP4AI), nine policy recommendations could be derived. Table 1 provides a systematic overview of these recommendations. The overview adapts the policy ontology originally developed by popAI [14], identifying for each recommendation at what level (Societal, Regulatory, Organisational, or Research) a recommendation should be implemented, whether the recommendation is reactively (📄) targeting the current state-of-play or proactively (🚀) anticipating new policy actions, who is the target audience for the recommendation, and which themes / aims are addressed by the recommendation. The recommendations are then described in more detail in the remainder of the section.

The ALIGNER project team graciously acknowledges that parts of these recommendations and their detailed descriptions were first published and updated by colleagues from the popAI project in [14, 15], while the initial ALIGNER policy recommendations were first published in September 2022 as part of ALIGNER D2.3 [15]. The ALIGNER and popAI project teams have since worked together to harmonize their recommendations. They presented these harmonized recommendations for the first time at a joint ethics event co-organized between DG Home, ALIGNER, AP4AI, popAI, and STARLIGHT in January 2023. The ALIGNER team has now iterated and extended these recommendations again – especially considering the AI Act – for publication in the roadmap²².

²² Please note that the recommendations provided in the following sections differ from the policy recommendations provided in ALIGNER deliverable D2.5, to ensure compatibility within the EU AI Cluster. As a result, some recommendations have adjusted wording or were merged with recommendations previously identified with the cluster projects. For the original ALIGNER recommendations, please see deliverable D2.5.



5.2.1 Recommendation overview

No.	Recommendation	Implementa- tion Levels	Type	Target audiences	Themes / Aims
1*	Ensure a constructive partnership between the AI Office and Member States' P&LEAs to ensure prevention of compliance issues, identification and exchange of best practices / lessons learned, and facilitate the joint co-creation of targeted guidelines for the implementation of the AI Act at P&LEAs.	Regulatory, Organisational		European Commission, AI Office, Member States, P&LEAs	Fairness, Transparency, Equality, Privacy, Human Rights, Non-Discrimination, Minimize misuse, Trustworthy AI, AI Applicability
2*	Explore the use of the EU Database for High-Risk AI systems by P&LEAs to facilitate exchange between European P&LEAs about the development, deployment, and use of High-Risk AI systems in compliance with the AI Act.	Regulatory, Organisational		European Commission, AI Office, Member States, P&LEAs	AI Applicability
3*	Clarify the meaning of “a genuine and present of foreseeable threat of a terrorist attack” in Article 5 of the AI Act to ensure ethical and legal use of remote biometric identification by P&LEAs.	Regulatory		European Commission, AI Office, EC DG Home	Fairness, Transparency, Equality, Privacy, Human Rights, Non-Discrimination, Minimize misuse, AI Applicability
4^	Embed the concepts of ‘AI Literacy’ and ‘human-centric approach’ into EU P&LEA training, including education on impacts, consequences, and implications of AI system use as well as use of real-world data for model training.	Regulatory, Organisational, Societal		European Commission, AI Office, DG Home, Member State Parliaments, Ministries, Europol, CEPOL P&LEAs	Fairness, Transparency, Equality, Privacy, Human Rights, Non-Discrimination, Minimize misuse, AI Applicability
5^	Establish and improve unified frameworks, compliant with the AI Act, for the evaluation of AI systems and models during development and deployment ensuring their ethical, legal, and societal compliance.	Regulatory, Organisational, Research, Societal		EC DG Home, EU Parliament, European Commission, Research Institutes, Industry, P&LEAs	Fairness, Transparency, Equality, Privacy, Human Rights, Non-Discrimination, Trustworthy AI
6^	Review existing and establish new legal and regulatory mechanisms to ensure that AI systems and their use are ethical, legal, and societally acceptable.	Regulatory		EU Parliament, European Commission, Member States Parliaments	Fairness, Transparency, Equality, Diversity, Social Inclusion, Privacy, Human Rights, Non-Discrimination, Minimize misuse, Trustworthy AI, AI Applicability






No.	Recommendation	Implementation Levels	Type	Target audiences	Themes / Aims
7 [^]	Develop meaningful dialogue between regulators, P&LEAs, researchers, industry, and civil society organizations to strengthen citizens' confidence in the use of AI systems by P&LEAs via the consultation processes of the AI Office and other means.	Regulatory, Organisational, Research, Societal		Member States, Parliaments, Ministries, P&LEAs, Research Institutes, Industry, Civil Society Organisations	Diversity, Transparency, Social Inclusion, Awareness, Trustworthy AI
8 [°]	Enable EU citizens to access basic information about AI systems used by P&LEAs.	Regulatory, Organisational		European Commission, AI Office, Member States	Fairness, Transparency, Equality, Privacy, Human Rights, Non-Discrimination, Minimize misuse, Trustworthy AI
9	Extend and adapt European and national research programmes to better facilitate evidence-based, participatory research into P&LEA needs regarding AI, the potential implications of the use of AI by P&LEA, and potential criminal use of AI.	Research		European Commission, Ministries / National Funding Agencies, Research Institutes, Civil Society Organisations	Social Inclusion, Trustworthy AI, AI Applicability

Table 1: Overview of policy recommendations. * = New recommendation; ^= revised recommendation; ° = adapted from popAI policy brief no. 2

5.2.2 Recommendations in detail

Recommendation 1

Ensure a constructive partnership between the AI Office and Member States' P&LEAs to ensure prevention of compliance issues, identification and exchange of best practices / lessons learned, and facilitate the joint co-creation of targeted guidelines for the implementation of the AI Act at P&LEAs.

The AI Act mandates the European Commission to "create guidelines on the practical implementation of this Act" and, upon request from Member States or the AI Office, or at its own discretion, "... update previously adopted guidelines as necessary."²³ One function of the AI Office is to aid the European Commission in preparing supportive guidance and guidelines for the Regulation's practical application. This includes developing tools such as standardised protocols and best practices in collaboration with relevant Commission bodies and agencies. Additionally, the AI Office is responsible for "promoting and facilitating the creation of codes of practice and conduct at the Union level,"²⁴ considering international approaches and monitoring the implementation and assessment of these codes. P&LEAs should

²³ AI Act, Article 96: Guidelines from the Commission on the Implementation of this Regulation

²⁴ Commission Decision Article 3(2)(i)



actively participate in drafting new codes of practice for AI systems that affect their roles and responsibilities.

The AI Office's involvement with the codes of practice specifies that *"The Commission may formalise a code of practice through implementing acts" or "it may establish common implementation rules through these acts."* Collaboration between the AI Office and EU P&LEAs on drafting these acts is advised to provide guidance to the Commission. This duty under the AI Act will likely be ongoing, beyond the initial implementation period.

The AI Act allows the AI Office to invite other entities to contribute to the development of codes of practice, supporting the AI Office. These could include *"civil society organisations, industry, academia, and other relevant stakeholders such as downstream providers and independent experts."*²⁵. According to Article 4 of the Commission Decision, the AI Office must consult stakeholders regularly, including experts from the scientific community and education sector, citizens, civil society, and social partners, to gather input for its tasks. Recognising the EU P&LEA community as a pivotal stakeholder, early consultations regarding codes of practice and potential guidance would benefit both parties. The AI Office would gain valuable insights from EU P&LEAs on the impact of the AI Act's compliance requirements. Furthermore, early dialogue between the AI Office and P&LEA representatives about implementation issues, legislation, guidance, best practices, and codes of practice would be mutually advantageous.

Considering the brief timeframe for drafting the codes of practice, the EU Innovation Hub at Europol could initially serve as a liaison between the AI Office and the national P&LEAs of each Member State until a more permanent arrangement is established. While Frontex (for border guards), Eurojust (for public prosecutors), and CEPOL (for law enforcement training) already engage with the Innovation hub, individual arrangements for these institutions could be explored as well.

The AI Office is also responsible for establishing a governance structure for the AI Act, coordinating an effective system, and setting up advisory bodies at the Union level. Accordingly, the AI Act proposes the creation of an AI Board composed of Member States' representatives, a Scientific Panel to integrate the scientific community, and an Advisory Forum for stakeholder input on the Act's implementation at both Union and national levels. Establishing a dynamic relationship between the AI Office and the EU P&LEA community can channel their insights to the Advisory Board, Forum, and sub-groups, offering unique information beyond the codes of practice or compliance needs. By doing this, the Commission would benefit from practical experience insights when applying the AI Act's provisions to real-world AI systems. Initially, Europol could be assigned as the first point of contact between national P&LEAs of Member States and the AI Office concerning AI system operations, until more permanent arrangements are confirmed.

It is anticipated that sustained communication from EU P&LEAs to the AI Office and additional entities will be crucial for staying updated with developments under the AI Act and addressing evolving challenges. Thus, extending communication to encompass broader knowledge and experiences from real-world AI system operations can help the AI Office contextualise and address legislative considerations due to its role in monitoring and regulating the AI Act's implementation.

²⁵ AI Act, Article 56: Codes of Practice, para. 3



Recommendation 2

Explore the use of the EU Database for High-Risk AI systems by P&LEAs to facilitate exchange between European P&LEAs about the development, deployment, and use of High-Risk AI systems in compliance with the AI Act.

The AI Act has a provision under Chapter VIII, Article 71 for the creation of an EU database of ‘High-Risk’ AI systems that are listed under Annex III of the Act. The types of system listed of most relevance to EU P&LEAs are those used for:

- ◆ Law enforcement
- ◆ Migration
- ◆ Biometrics

Annex VIII, Section C – Information to be submitted by deployers of high-risk AI systems in accordance with Article 49(3), requires the deployer to provide and keep updated the following information:

- ◆ The name, address and contact details of the deployer
- ◆ The name, address and contact details of the person submitting information on behalf of the deployer
- ◆ The URL of the entry of the AI system in the EU database by its provider
- ◆ A summary of the findings of the fundamental rights impact assessment conducted in accordance with Article 27
- ◆ A summary of the data protection impact assessment carried out in accordance with Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680 as specified in Article 26(8) of this Regulation, where applicable

A mechanism should be established to enable the information detailed above to be released to an appropriate employee of a P&LEA for purposes of researching their organisations capability needs, potential procurement process, development or operational issues relating to high-risk AI systems.

Article 71(4) says that *“the information registered in accordance with Article 60 shall be accessible only to market surveillance authorities and the Commission, unless the prospective provider or provider has given consent for also making the information accessible the public.”* However, an argument can be made for the Commission to set up a specific section of the register allowing EU P&LEAs to voluntarily share with selected third parties (e.g., with other EU P&LEAs only) what AI systems they operate, so to facilitate communications on experiences in procurement, operations, etc.

Recommendation 3

Clarify the meaning of “a genuine and present of foreseeable threat of a terrorist attack” in Article 5 of the AI Act to ensure ethical and legal use of remote biometric identification by P&LEAs.

The phrasing of Article 5 in the AI Act mandates that P&LEAs must prove the necessity of employing a real-time remote biometric identification system to prevent a *“genuine and present or genuine and foreseeable threat of a terrorist attack.”* They must comply with all other stipulations of the Act as well.



However, the terms ‘genuine,’ ‘present,’ and ‘foreseeable’ used in the Act to describe the type of terrorist threat are problematic due to their ambiguity, subjectivity, and open-ended nature. These terms are not commonly found in established terrorist threat assessment methods, making them unsuitable for P&LEAs to use in demonstrating compliance with the AI Act under such circumstances.

The goal of any terrorist threat assessment is to identify necessary preventive measures based on information gathered and evaluated before an attack occurs. This information must be assessed for reliability both in terms of its source – graded from ‘Reliable’ (Grade A) to ‘Cannot be judged’ (Grade F) – and the information itself – from ‘Confirmed’ (Grade 1) to ‘Cannot be judged’ (Grade 6). Once reliability is established, the specific threat assessment hinges on two critical factors: The credibility of the terrorist threat and, if the threat is credible, the likelihood of it materializing.

In security threat assessments, a ‘credible’ threat indicates that the adversary has both the intent and the capability to execute the threat. If either intent or capability is lacking, the threat may exist but at a diminished level. The ‘probability’ or likelihood of a threat can be classified on a scale ranging from ‘Rare or remote,’ through ‘Unlikely,’ ‘Credible,’ and ‘Likely,’ to ‘Almost certain,’ each with defined meanings in the context of terrorist threats and security responses.

If the need for criteria such as those described above is endorsed by the AI Office as appropriate for P&LEAs, there will be a requirement to standardise the degrees of ‘credibility’ and ‘probability’ that P&LEAs must demonstrate to legally deploy such AI systems in response to terrorist threats.

Recommendation 4

Embed the concepts of ‘AI Literacy’ and ‘human-centric approach’ into EU P&LEA training, including guidance and education on impacts, consequences, and implications of AI system use as well as use of real-world data for model training.

Interactions during multiple activities of the EU AI Cluster comprised of ALIGNER, AP4AI, popAI, and STARLIGHT, including exchanges with other projects (see Annex A), survey results (see section 3 and Annex B), as well as other research activities [16] highlight the need for and the lack of clear guidelines, education, and training for P&LEAs regarding issues surrounding the development, procurement, deployment, and use of AI systems. This is in line with the requirements from the AI Act towards providers and deployers of AI systems to ensure a sufficient level of AI literacy, i.e. *“skills, knowledge and understanding that allows providers, users and affected persons, taking into account their respective rights and obligations in the context of this [AI Act] Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause.”*²⁶

As such, P&LEAs as users of AI systems require education and training on the correct, safe, and responsible use of individual AI systems. This also includes, even more importantly, guidance and training on the reliable evaluation of the ethical, legal, and societal implications of the use of AI (see also recommendation 5), supporting effectiveness of AI evaluations by moving away from a black box approach towards explainable AI. This should also cover the necessary knowledge and (regular) training about what to ask for (i.e., the requirements given by the legal and ethical considerations) and

²⁶ AI Act, Article 3(42)(bh)



what to expect from the providers of AI systems, as well as the capability to regularly and systematically monitor the ethically, legally, and socially acceptable operation of such systems based on the instructions of use.

A specific issue in the development and deployment of AI relates to data protection and the necessary trade-off between protecting personal, sensitive data and the need for large ‘real-world’ datasets for training applicable AI models. Specific guidance and training on how to ensure data protection while simultaneously allowing for training AI models with real-world applicability is very much needed.

To support the systematic embedding of AI literacy into P&LEA training a structured and agile and life-long learning educational approach is essential. This approach needs to be complemented by specific courses, e.g. on “ethical and law AI”, part of the national educational programmes of LEAs during their studies at the police academies. With the support of Europol, CEPOL, and the AI Office, relevant seminars, courses, and workshops on the European level could be established.

Recommendation 5

Establish and improve unified frameworks, compliant with the AI Act, for the evaluation of AI systems and models during development and deployment ensuring their ethical, legal, and societal compliance.

The guidelines and support needed to ensure ethical, legal, and societal compliance, as well as the actual applicability of AI systems, need to be grounded in evidence-based, unified evaluation frameworks. Given the special role of P&LEAs within society, such assessment frameworks will need to follow a broader approach to impact assessment. As identified by popAI, the literature proposes several AI system assessment frameworks^{27,28,29,30} as well as methods that provide indicators of risks a company might face when adopting an AI system, while also including mitigation actions and best practices that might be followed. Each of these frameworks includes different guidelines, assessment criteria and mitigation recommendations concerning the adoption of AI. However, most of them focus on the private sector, resulting in a lack of assessment frameworks and clear implementation procedures that provide guidelines, recommendations, and mitigation indicators that can support AI literacy in the public sector (see also recommendation 4). The AP4AI Framework for assessing the accountability of AI systems as well as the ALIGNER Fundamental Rights Impact Assessment [17] (which is based on the MAGNETO³¹ Ethical Risk Assessment Form) take steps in this direction but need to be further aligned with other frameworks.

Therefore, there is both an ongoing need for more extensive research on the development of such frameworks and interdisciplinary assessment measures/metrics as well as relevant standardisation efforts on the European and international level. The latter is already ongoing, e.g. with the establishment of the CEN-CENELEC Joint Technical Committee 21 on “Artificial Intelligence”³².

²⁷ High-Level Expert Group (HLEG) - Assessment List for Trustworthy Artificial Intelligence (ALTAI): <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

²⁸ World Economic Forum (WEF) AI Governance framework: <https://www.weforum.org/projects/model-ai-governance-framework>

²⁹ NOREA Guiding Principles Trustworthy AI investigation: <https://www.norea.nl/uploads/bfile/a344c98a-e334-4cf8-87c4-1b45da3d9bc1>

³⁰ AI Assessment Catalogue of Fraunhofer IAIS: <https://www.iais.fraunhofer.de/en/research/artificial-intelligence/ai-assessment-catalog.html>

³¹ <https://www.magneto-h2020.eu/>

³² <https://www.cenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/>



Recommendation 6

Review existing and establish new legal and regulatory mechanisms to ensure that AI systems and their use are ethical, legal, and societally acceptable.

Operative guidelines for the development, procurement, deployment, and use of AI systems and models, based on evidence-based, unified evaluation frameworks, will need to be flanked by binding legal mechanisms to ensure that these technologies are ethical, legal, and societally acceptable. The AI Act is a step in this direction, although based on numerous discussions with representatives from P&LEA, civil society, research, industry, and policy, there remain valid concerns from different actors on its definition of AI (too broad), the exemptions included for high-risk AI technologies (too many), and its affect when put into place (too bureaucratic). A valid approach to alleviate these concerns might be the development of a P&LEA-specific AI directive (similar to the Law Enforcement Directive [18]).

The review of existing and establishment of new legal and regulatory mechanisms should specifically include the promotion of inclusiveness and gender diversity in AI for P&LEAs, as this is a critical factor for establishing trust and ensuring equitable and effective use of AI systems. This could be achieved by the establishment of inclusive AI development standards and frameworks for police and law enforcement to guide P&LEAs, including:

- ◆ Ensuring diverse AI development teams that are composed of members with different genders, ethnicities, and expertise in ethics and bias mitigation;
- ◆ Ensuring a transparent documentation of the AI development process, employed data sources, algorithms, and models – as requested by the AI Act; and
- ◆ Ensuring regular third-party audits of AI systems in use by P&LEAs to identify and resolve bias and fairness issues.

Recommendation 7

Develop meaningful dialogue between regulators, P&LEAs, researchers, industry, and civil society organizations to strengthen citizens' confidence in the use of AI tools by P&LEAs via the consultation processes of the AI Office and other means.

Civil society organisations are often not included in consultations regarding the employment of AI systems by P&LEAs. Therefore, they express their concerns on emerging risks through announcements and legal actions. This gap is creating tensions that are constantly widening and damage the trust between the involved parties.

To repair the trust issues, civil society organisations should be explicitly involved in open dialogues with European and national regulators, P&LEAs, researchers, and industry regarding the employment of AI systems via the consultation processes of the AI Office mandated by the AI Act. In this way, civil society organisations could be actively involved in the process of designing and implementing AI systems, as well as in the monitoring of the existing ones. They should also determine the best way to operate these systems to ensure human rights and generate acceptance across citizens.

Another way to better integrate civil society organisations and foster dialogue would be to create joint national / local working groups, which could check the individual AI systems used by P&LEAs to



highlight potential issues from such usage (a posteriori monitoring and assessment). These joint working groups should also be consulted when designing and developing new AI systems that will be applied in the future (a priori monitoring and assessment). The purpose is to improve and adapt these systems appropriately to ensure that they protect citizens' rights. This interaction between different actors related to the use of AI technologies by P&LEAs should be continuous and should strengthen the involvement of civil society in all stages of the operation of an AI system (design, implementation, maintenance, upgrade; see also recommendation 9). To facilitate this interaction, the European Commission and EU Member States need to better promote and ensure citizens' awareness regarding the existence and implementation of an AI system and enable objection to potential unjust decisions (see recommendation 8).

Open discussions between different actors related to the use of AI systems by P&LEAs can support transparency at every stage to minimize the risks of discrimination. In addition, this should also be considered in the procurement of systems, where, for example, the technical specifications must be accepted by civil society organizations and agencies, while monitoring and assessment by representatives of social and other bodies should be foreseen in the system implementation phase.

Recommendation 8

Enable EU citizens to access basic information about AI systems used by P&LEAs.

Citizens' concerns about the use of AI systems by LEAs, combined with the real ethical and legal risks of their use, make it necessary for all stakeholders to know what is actually there and what is "at risk"³³. As such it is highly recommended that the Commission and the AI Office explore the extension of (part of) the EU database of 'High-Risk' AI systems, established under Chapter VIII, Article 71 of the AI Act, as a means to provide basic information about each AI system used by each LEA at a country level to all EU citizens.

The information to be included in the database should cover a brief description of the AI system use, a general – layman's – description of the technology it uses, when it was designed/implemented, its data sources or which type of data is being used, when it was deployed, if and when and by whom it has undergone GDPR compliance checks, impact assessments, etc. Access to all or part of this information may be available to all citizens or only to interested parties after request to balance the needs of public security with the rightful concerns regarding individual rights (see also research recommendation 5). In this way, transparency will be enhanced, control will be strengthened, and a uniform approach to the legal and ethical use of AI by P&LEAs will be achieved. Since transparency is a cornerstone of trust, such a use of the database would ensure citizens trust in AI and at the same time it would ensure the responsible AI use, and promotion of collaboration among agencies within Europe.

Recommendation 9

Extend and adapt European and national research programmes to better facilitate evidence-based, participatory research into P&LEA needs regarding AI, the potential implications of the use of AI by P&LEA, and potential criminal use of AI.

³³ This recommendation was first published by the popAI project in its second policy brief. It has been adapted here in light of the provisions of the AI Act.



EU- and nationally funded security projects, and specifically those developing AI driven technologies, have often raised concerns, see for example the FP7 project INDECT “Intelligent information system supporting observation, searching and detection for security of citizens in urban environment”³⁴, which sparked concerns among Members of European Parliament calling on the European Commission to clarify its purpose³⁵. The – sometimes overly restrictive – secrecy of such projects and lack of publicly available information, together with the perceived potentially negative impact on civil liberties and fundamental rights call for new approaches towards accountability. One way to address these issues, while maintaining the required level of security, would be the establishment of specialised interdisciplinary Ethics and Legal Committees – potentially located at the AI Office – who review proposals and ongoing research projects in the security domain on a continuous basis, to prevent potentially serious ethical, societal, and legal issues as well as abuse of human rights. Aligned with recommendations 4 and 5 these Committees should have ethical, legal, technical, organisational, and practical capabilities to assess an AI systems’ ethical, legal, and societal compliance. This could act as a form of internal certification for research projects in relation to an AI systems’ accountability and the ethical, inclusive and secure-by-design AI systems during research and development.

In addition, research conducted in the context of the H2020 project popAI identified the stakeholder groups involved in the research, development, use, and implementation of AI technology, as well as those who promote awareness regarding emerging risks, and push for relevant policies. These different categories of stakeholders should not be seen as “rivals” but rather as key components of a unified ecosystem that co-shape the development and use of AI in the security domain. The identified stakeholders are namely, LEAs, social and humanities research, policy makers, government and public bodies, technologists / data scientists, civil society organizations, national and local authorities, ICT and software companies, and police academies. Mapping EU-funded projects in the security domain, 348 different stakeholders were collated with most stakeholders being ICT and software companies, followed by universities and research organisations. It is recommended that the EC explores ways (i.e., call requirements, specifications) for EU-funded projects to include civil society organisations in the early stages of the AI technology design and development as they are underrepresented in the project consortia, while their voices are very important to preserve privacy and human rights. Likewise, project partners were geographically mapped. The analysis indicated that various European countries such as Albania, Denmark, and Ukraine have been underrepresented to date in EU-funded projects in the security domain. Involvement of partners from underrepresented European countries would enable the inclusion of potentially cultural and geographic differences regarding the needs and acceptance of AI systems. Thus, it is recommended that the EC explores ways (i.e., call requirements, specifications) for EU-funded projects to include underrepresented Member States in the AI design and development.³⁶

Lastly, the implementation of recommendations 1-8 needs to be supported by further AI-specific research in the security domain. This includes the development of guidelines aligned with the needs of P&LEAs (recommendation 1, 3, 4, 5), assessment frameworks (recommendation 5), an evaluation of the existing legal mechanism as well as their effects on P&LEA work (recommendation 6), stakeholder engagement techniques in the context of AI systems for P&LEAs (recommendation 7), as well as guidelines for ensuring inclusivity and diversity when developing AI systems (recommendation 6). This

³⁴ INDECT (Intelligent Information System Supporting Observation, Searching and Detection for Security of Citizens in Urban Environment), Cordis Project Page.

³⁵ Euractiv (2011), “MEPs question ‘Big Brother’ urban observation project”.

³⁶ This paragraph was first published by popAI in [13].



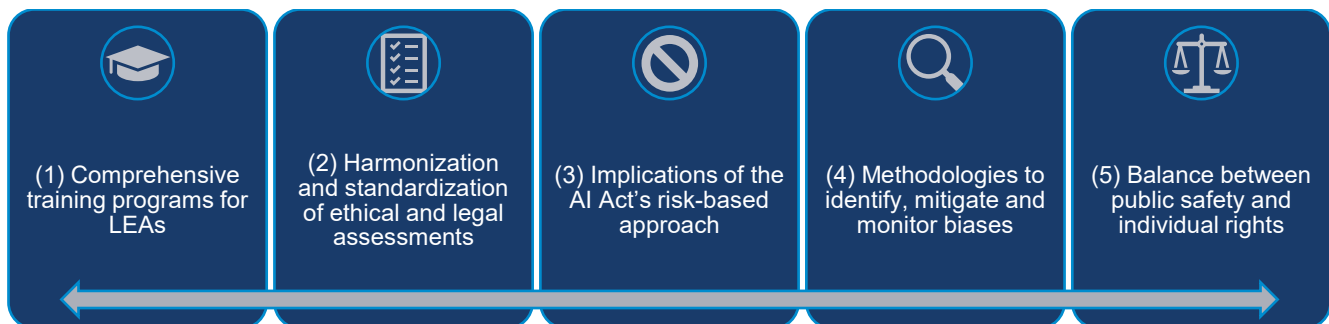
also includes additional research into countering criminal use of AI technologies and employing AI systems in support of P&LEAs in an ethical, legal, and societally acceptable way.

5.3 Research recommendations

Using the capability enhancement needs and trends in the use of AI technologies identified in surveys and workshops, ALIGNER identified the most pressing research needs in the realm of AI used by P&LEAs. These needs were further validated through specific discussions with members of ALIGNER's advisory boards and using insights from other relevant events involving researchers and P&LEA practitioners. These needs have been methodically categorized into ethical, legal, societal, and technical aspects and further divided into short- to medium-term and medium- to long-term requirements, taking the usual development processes of European research framework programmes into account, i.e. short- to medium-term research needs express urgent requirements given the current unprecedented rise in the use of AI and should be addressed within the next 2-5 years, while medium- to long-term research needs address complex problems that might not be able to be addressed immediately and might be based on results from short-term research needs.

5.3.1 Short- to medium-term research

5.3.1.1 Ethical, legal, and societal aspects



(1) There is an urgent need to develop **comprehensive training programs for Law Enforcement Agencies** on the ethical and legal implications of AI systems (see also policy recommendation 4). These programs should aim to educate officers and other stakeholders about the principles of human oversight, fairness, accountability, and transparency in AI usage. The primary focus of such training should be to ensure that everyone involved in the deployment and operation of AI systems understands the ethical and legal implications of the deployment of AI for law enforcement purposes. This includes recognizing and mitigating biases in AI algorithms, ensuring that the systems are used in a manner that respects fundamental rights, and maintaining transparency about how AI decisions are made. Furthermore, training should cover the frameworks and regulations governing AI systems, such as the AI Act and the Law Enforcement Directive. Officers need to be well-versed in these laws to ensure compliance and to understand the legal boundaries within which AI can be used. This knowledge is crucial to avoid the misuse of AI and to protect individuals' rights.

(2) Efforts should also be directed towards the **harmonization and standardization of ethical and legal AI assessments** on national, European, and international scales. This involves developing standardized protocols and evaluation metrics that can be universally applied to ensure consistency and reliability in AI assessments. By having a common set of standards, it becomes easier to compare,



trust, and ensure that AI systems are used equally responsibly and fairly worldwide. This is particularly important as AI systems are increasingly used in law enforcement, where they can have significant impacts on people's lives, also beyond one country's jurisdiction. Standardization helps to avoid discrepancies and ensures that the same high standards are applied everywhere, thereby promoting fairness, accountability, and transparency in AI usage.

(3) With legislative frameworks such as the AI Act coming into effect, it's essential to **comprehend its implications, particularly regarding which AI systems are categorized as prohibited or high-risk.** The current categorization is based on the potential for significant harm to fundamental rights, safety, or societal values, but the criteria for determining what constitutes a "prohibited" or "high-risk" system require further refinement and empirical validation. Harmonized guidance, based on empirical studies, is needed to further precise which AI systems are to be considered as prohibited, especially in the case of predictive policing systems. Moreover, research is needed to develop more quantifiable metrics and clear definitions for assessing risk levels, considering factors such as the context in which the AI system is deployed, the sensitivity of the data it processes, and its potential societal impact, especially to harmonize the implementation of Article 6(3) of the AI Act.

One key area of study is the development of risk assessment frameworks that can evaluate not only technical risks, such as accuracy and robustness, but also broader ethical and social implications, such as bias, fairness, and privacy and data protection violations. Researchers should investigate how to create adaptable, context-aware models that can evaluate risk dynamically as AI systems evolve. Additionally, comparative research on the intersection of risk and sector is needed, as the implications of using AI in healthcare differ significantly from those in law enforcement or financial services. This research could inform more nuanced regulatory guidelines, ensuring that the categorization process is both effective and flexible enough to account for future AI innovations.

Furthermore, the AI Act requires continuous monitoring and reporting to ensure compliance, which raises challenges in operationalizing these mandates in real-world systems. Researchers must focus on creating tools and frameworks for continuous risk assessment, auditing AI models, and ensuring human oversight in decision-making.

(4) AI systems are susceptible to **biases that can lead to unfair treatment of individuals.** Research should focus on developing methodologies to identify, mitigate, and monitor bias in AI decision-making processes. This includes establishing baseline data on human bias and exploring how AI can be less biased, which will also allow to identify contexts in which AI systems can reduce bias and support research into validating and cleansing historical datasets (see recommendations 14 and 15).

Bias in AI systems can stem from the data they are trained on, which may reflect existing prejudices or inequalities in society. For example, if an AI system is trained on historical data in which certain groups were underrepresented or treated unfairly, it might perpetuate those biases in its decisions. To counteract this, researchers need to develop methods that can detect and address bias at various stages of the AI lifecycle – from data collection and preprocessing to model training and deployment.

Furthermore, the design of the algorithms themselves can introduce bias. Algorithms are created based on human decisions about what factors and variables are important, which can be influenced by the designers' own biases and assumptions. For instance, if an algorithm is designed to prioritize certain criteria over others without considering the broader context, it might inadvertently favour one group over another. This can happen if the criteria chosen for decision-making reflect societal biases, leading to algorithmic decisions that reinforce existing inequities. To mitigate this, it is crucial to incorporate diverse



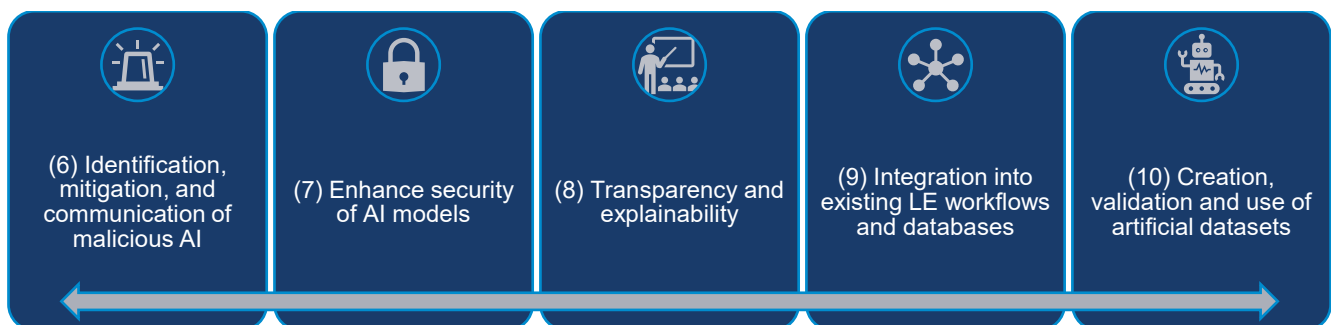
perspectives during the design process (see also policy recommendation 6). This includes involving individuals from different backgrounds and fields to ensure that a wide range of viewpoints and potential biases are considered. Additionally, employing techniques such as fairness constraints and bias audits can help identify and correct biases in the algorithm before it is deployed. Continuous monitoring and iterative testing are also essential to ensure that the algorithm remains fair and unbiased over time.

Another approach is to implement fairness-aware algorithms [19] that can adjust for biases in the data. These algorithms can be designed to ensure that the AI system's decisions do not disproportionately impact any particular group. Additionally, continuous monitoring and auditing of AI systems are essential to identify and correct any biases that may emerge over time. Moreover, it is important to conduct thorough testing and validation of AI systems using diverse datasets that represent different population groups.

(5) A critical area of research is finding the **balance between ensuring public safety and protecting individual rights**. This includes developing privacy-preserving AI methods that can safeguard citizens' personal information while still enabling effective law enforcement (see also recommendations 7, 10, and 19). For instance, techniques such as differential privacy [20] and federated learning [22] can be employed to anonymize data, thus ensuring that individual identities are protected. Additionally, researchers should explore encryption methods that allow AI systems to process sensitive information without exposing it. It's essential that these methods be robust and adaptable to different contexts to maintain a high level of security.

Moreover, it involves creating transparent and accountable AI systems that law enforcement agencies can use without infringing on civil liberties. This requires developing algorithms that are explainable, meaning that their decision-making processes can be easily understood and scrutinized by human operators. Establishing clear protocols for human oversight and intervention is also imperative to prevent misuse and ensure that AI decisions are always subject to human review.

5.3.1.2 Technical aspects



(6) As AI becomes more prevalent, the **identification, mitigation, and communication of malicious AI models** become crucial. Malicious AI models can be designed to harm users, manipulate data, or execute unfair or biased decisions. To prevent these adverse outcomes, research should focus on developing techniques to detect and neutralize such threats. This involves creating algorithms that can identify unusual or harmful behaviour in AI models, implementing robust security measures to protect against attacks, and continuously monitoring AI systems for any signs of compromise. It is equally important to communicate instances of malicious AI models to users and sensitize them to the potential risks associated with such models. These efforts will ensure the safe deployment of AI in various fields,



including law enforcement, where the risks associated with malicious AI can have significant impacts on public safety and trust.

(7) There is a pressing need to **enhance the security of AI models** to ensure that the data used for training these models is not identified or misused. To achieve this, research should delve into advanced encryption techniques, which can encode data in such a way that only authorized parties can access it. Additionally, data anonymization methods should be explored. These methods strip personal identifiers from datasets, making it impossible to trace the data back to individual sources. By implementing these security measures, we can protect sensitive information and foster greater trust in AI systems. Furthermore, these techniques will also aid in preventing malicious actors from exploiting AI models for harmful purposes. By reinforcing the security of AI models, we ensure their safe application across various fields, including law enforcement, healthcare, and finance, thereby safeguarding public welfare and maintaining the integrity of AI-driven decision-making processes. The security of AI models (and the associated data) is also a prerequisite for subsequent research on methods to allow continuous data exchange between different AI systems – potentially across borders (see recommendations 17).

(8) For AI systems to be effectively integrated into law enforcement workflows, they must be interpretable and transparent. This means that the AI systems should be able to provide clear and understandable explanations for their decisions. When AI systems can explain how they arrive at certain conclusions, it not only increases the trust of human operators but also ensures that the decisions made are fair and justified. Research should therefore focus on creating **AI systems that can offer detailed insights into their decision-making processes**, thus enhancing their overall accuracy and reliability.

(9) Research should address **how AI systems can be better integrated into existing law enforcement workflows and databases**. This includes developing standardized APIs (Application Programming Interfaces) that allow different software systems to communicate with each other efficiently. Standardized APIs ensure that AI systems can be easily added to current systems without requiring extensive modifications, thereby facilitating a smoother implementation process. Additionally, using standardized data formats can help ensure consistency and compatibility between AI systems and existing databases. This standardization makes it easier to exchange and interpret data across various platforms, which is crucial for accurate and reliable AI-driven decision-making.

Moreover, hybrid cloud-based solutions, such as the European Data Spaces³⁷, provide a flexible and scalable infrastructure for integrating AI systems. These solutions combine on-premises resources with cloud services, allowing law enforcement agencies to benefit from the scalability and computational power of the cloud while maintaining control over sensitive data. Hybrid cloud architectures enable agencies to deploy AI systems more effectively and ensure that the integration process does not disrupt their existing operations.

(10) To train robust AI models, there is a need for high-quality artificial datasets. Research should explore the **creation, validation, and utilization of artificial datasets** that can effectively simulate real-world scenarios. These datasets can be generated using various techniques such as data augmentation, synthetic data generation, and simulation environments. Data augmentation involves creating new data points from existing data through transformations such as rotation, scaling, and flipping. This technique enhances the diversity of the training data, making AI models more resilient to

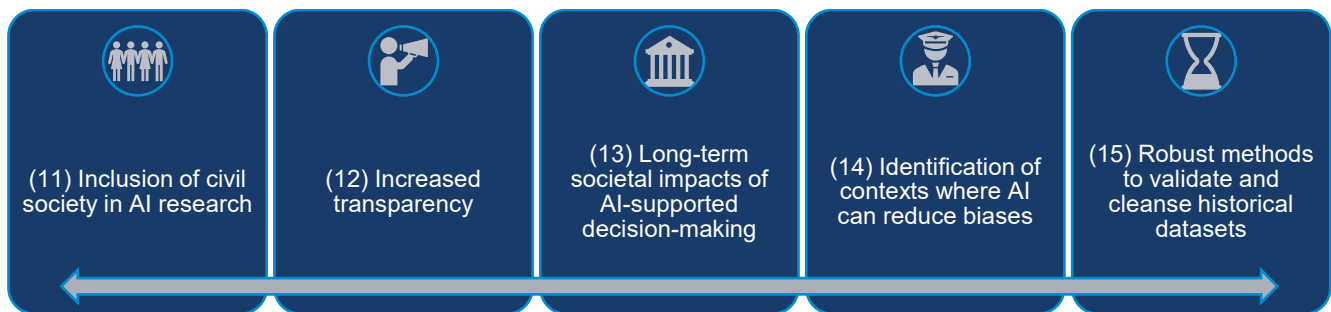
³⁷ <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>



variations in real-world inputs. Synthetic data generation creates entirely new datasets using algorithms that mimic real data's statistical properties. This approach can be particularly useful when real data is scarce, sensitive, or expensive to obtain (see also recommendation 5). However, when creating synthetic datasets special care needs to be taken to not perpetuate biases (see recommendation 4). Simulation environments allow researchers to create controlled, virtual scenarios where AI models can be trained and tested. These environments can replicate complex situations that may be difficult or risky to reproduce in the real world, such as emergency responses or autonomous vehicle navigation.

5.3.2 Medium- to long-term research

5.3.2.1 Ethical, legal, and societal aspects



(11 & 12) Long-term research should focus on **developing better methods to include civil society in AI security research**. This is crucial to ensure that the deployment of AI systems is transparent, accountable, and aligned with public values and interests (see recommendations 5 and 8). To achieve this, several strategies can be implemented:

- ◆ **Public Engagement Strategies:** Establishing forums, workshops, and public consultations where citizens can voice their concerns and provide input on AI-related projects. This ensures that the public's perspective is considered in the development and deployment of AI systems.
- ◆ **Transparency Mechanisms:** Creating clear and accessible information about how AI systems function, their intended use, and their potential impacts. Transparency fosters trust and allows for informed public discourse.
- ◆ **Effective Communication Channels:** Developing robust channels of communication between AI researchers, developers, deployers, and the public. This could include regular updates, newsletters, and interactive platforms where people can ask questions and receive timely responses.
- ◆ **Inclusive Policy Development:** Involving a diverse range of stakeholders, including marginalized groups, in the policymaking process to ensure that AI systems are equitable and do not reinforce existing biases.

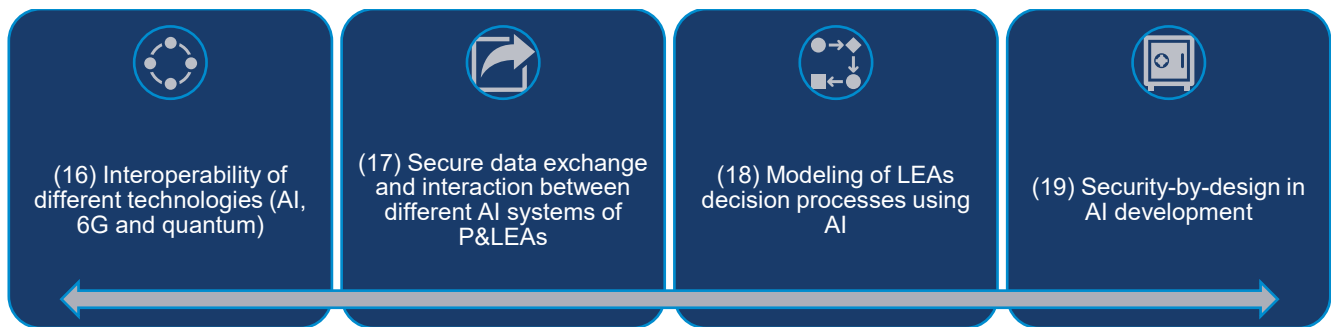
(13) Research should investigate the **long-term societal impacts of AI-supported decision-making in criminal investigations and in the criminal justice system**. This includes understanding how AI influences decision-making processes and the potential societal changes that may result from its widespread use. Specifically, it is crucial to explore how AI can affect the fairness and transparency of P&LEAs' and judicial decisions and how it may alter the dynamics of trust between the public and democratic institutions.



(14) Moreover, this research should delve into the **contexts in which AI can help reduce biases** inherent in human decision-making (see also recommendations 4). By analysing historical data and current practices, researchers can identify areas where AI can provide a more objective assessment and mitigate personal or systemic biases that have long plagued criminal investigations and the criminal justice system.

(15) Ensuring that historical data used in AI systems is complete and error-free is essential. Research should focus on **developing robust methods to validate and cleanse historical datasets**. This includes identifying and correcting inaccuracies, filling in missing information, and ensuring that the data accurately represents the context and conditions under which it was collected. High-quality data is crucial for training AI models that make fair and reliable decisions, particularly in sensitive areas like the criminal justice system. By improving the integrity of historical data, researchers can enhance the objectivity and effectiveness of AI systems, thereby fostering greater trust and transparency in AI-driven decision-making processes.

5.3.2.2 Technical aspects



(16) As technologies such as AI, 6G, and quantum computing converge, research should focus on their **interoperability**. This involves developing standards and protocols that enable these diverse technologies to communicate and work together seamlessly. By ensuring that AI systems can interact effectively with other emerging technologies, we can enhance their capabilities and create more comprehensive solutions.

Understanding and addressing the challenges of interoperability is crucial for the success of integrated technological ecosystems. For example, AI systems need to be able to process and analyse data from various sources, including future 6G networks that offer unprecedented speed and bandwidth. Similarly, quantum computing can provide new levels of computational power, enabling AI systems to solve complex problems more efficiently. However, this requires that these systems are designed to integrate and leverage each other's strengths.

Furthermore, interoperability is essential for the deployment of AI systems in critical areas such as law enforcement and public safety. By facilitating data exchange and interaction between different AI systems, such as those used by various law enforcement agencies, we can improve the effectiveness and coordination of their efforts. This can lead to more accurate and timely responses to incidents, better strategic planning, and enhanced overall security.

To achieve this level of interoperability, research should focus on several key areas:

- ◆ **Standardization:** Developing common frameworks and standards that ensure compatibility between different technologies.



- ◆ **Communication Protocols:** Creating robust and secure protocols for data exchange and interaction.
- ◆ **Integration Testing:** Conducting extensive testing to identify and resolve any issues that may arise when integrating multiple technologies.

(17) Research should explore ways to facilitate **secure data exchange and interaction between different AI systems for P&LEA use**, for example, through EU Data Spaces. This can enhance the capabilities of AI systems and promote more effective law enforcement strategies.

EU Data Spaces are frameworks established within the European Union to ensure the secure and efficient sharing of data across various sectors and platforms. By leveraging such frameworks, AI systems could securely access a broader spectrum of data sources, leading to more comprehensive and accurate analyses. This is particularly beneficial for AI systems for P&LEA use, as it enables them to gather and process data from multiple jurisdictions and databases, ensuring a more coordinated and unified approach to crime prevention and investigation.

Furthermore, facilitating data exchange through such standardized platforms ensures that the data being used is consistent, reliable, and up to date. This reduces the risk of discrepancies that could arise from using isolated datasets and enhances the overall robustness of AI-driven decision-making processes. By integrating data from diverse sources, AI systems can generate more nuanced insights, which are crucial for developing effective strategies and responses in real-time situations.

Additionally, the interoperability fostered by these data exchange frameworks supports more collaborative efforts between different P&LEAs and other relevant stakeholders. This collaboration is essential for tackling cross-border crimes and ensuring public safety on a larger scale. Enhanced secure data sharing not only improves the efficiency of operational activities but also contributes to more informed policy-making and strategic planning within the realm of law enforcement.

(18) Research should investigate **how to model the decision processes of P&LEAs using AI**. This includes using historical decision-making data to train AI models that can support strategic planning and operational decision-making. By analysing past decisions and outcomes, these models can identify patterns and provide insights into effective strategies, helping P&LEAs to make more informed and timely decisions in future scenarios. This modelling effort also involves understanding the context in which decisions were made, including the challenges faced and the resources available at the time, ensuring that the AI recommendations are both realistic and applicable to real-world situations. Additionally, such models can simulate various scenarios, allowing P&LEAs to test and refine their approaches before implementing them in the field, ultimately enhancing their preparedness and response capabilities.

(19) Long-term research should also focus on **incorporating security-by-design principles in AI development**. This involves ensuring that AI systems are designed and trained with security considerations from the outset to prevent misuse and abuse.

By integrating security measures during the initial stages of AI development, developers can create systems that are inherently more resilient to attacks and vulnerabilities. This means thinking about potential threats and designing AI systems in such a way that they can detect, prevent, and respond to security breaches effectively. For example, this could involve implementing robust authentication mechanisms, secure data handling practices, and regular security audits throughout the AI system's lifecycle.



Moreover, security-by-design also means that AI systems should be able to adapt to new and emerging threats over time. This requires continuous monitoring, updating, and re-training of AI models to ensure they remain secure against evolving cyber threats. By prioritizing security from the beginning, AI developers can help protect sensitive data, maintain user trust, and ensure that their technologies are used safely and responsibly.



6. AI Technology Catalogue

This section provides a detailed description and an assessment of the AI technologies with relevance for the ALIGNER narrative. For each technology, a brief description is provided. In addition, each entry provides information on

- ◆ **Effectiveness** – A rough estimate in the short-term on effectiveness and performance described in non-technical language.
- ◆ **Robustness** – An assessment in the short-term on how robust the technology is for being able to handle counter measures, data quality issues and out-of-distribution examples (examples of a type it has not been trained on).
- ◆ **Development** – A mid-term perspective of what the current development efforts are and who are doing it. A general assessment of where the technology is heading within the next few years.
- ◆ **Projected future** – Long-term perspective of where this technology may end up a few years from now.
- ◆ **TRL** – An assessment of maturity using the simplified Technology Readiness Level scale.
- ◆ **Additional information** – Additional information on general functionality, application area, data sources, and algorithms used for the technology.
- ◆ **Risks and mitigation measures** – An assessment of the potential ethical, legal, and technological risks associated with the technology and suggestions for suitable mitigation measures

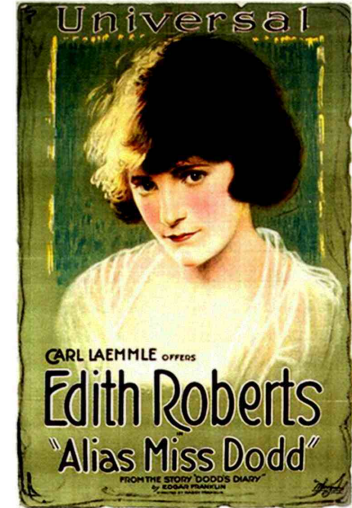
The assessment of effectiveness, robustness, development, and projected future uses admiralty code: confirmed, probably true, possibly true, doubtful, improbable, cannot be judged.



6.1 Deanonimization – Authorship Attribution

Authorship attribution is the task of identifying the author of a given text document within a set of possible candidates. A set of relevant textual features are used to create a "fingerprint" of the author. This "fingerprint" can be matched against a given set of candidates.

Examples of crimes where authorship attribution is important are illegal drug marketing, online threats, and extremism propaganda. Authorship attribution has shown promising results for e-mails, forum posts, tweets, and blog posts.



Picture from Wikimedia, Creative Commons 4.0

Effectiveness (short-term perspective): High

Can reliably find the matching author for a variety of textual content. However, the accuracy is not sufficient to use it as evidence in courts.

Robustness (short-term perspective): Medium

Can only match authors within a set of known candidates. Open authorship attribution where the author may not be among the candidates is much harder. Probably requires reasonably sized samples of written texts from all candidates.

Development (medium-term perspective): Active

Active commercial and academic research by many different actors. Recent efforts directly use source material to implicitly learn relevant features. This improves performance considerably compared to previous approaches.

Projected future (long-term perspective): Promising

Image databases from social media and satellites will increase over time. Future development will increase performance and have high precision for even larger geographic areas.

TRL: Currently 4-6

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Surveillance Data
Publicly Available Information

Algorithms:
Natural Language Processing,
Deep Learning

**Risks:**

- If the output of the AI system cannot be traced back and explained, due process rights may be undermined.
- To increase its accuracy, the AI system needs to be trained with a vast amount of data collected from various (online) sources, including e-mails, social media posts and blog posts. This may violate the right to data protection and, especially, the principle of lawfulness and data minimisation.
- Since it undermines the group anonymity, the AI system may have chilling effects on the freedom of expression.
- Depending on the envisaged use (e.g., in the context of extremism or terrorist propaganda), the AI system may be biased and systematically target members of certain social groups. This may violate the right to non-discrimination.

Mitigation measures:

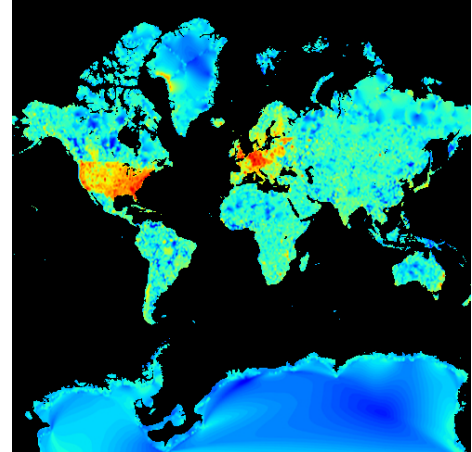
- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Ask AI developers to explain their algorithms' decision-making processes; implement procedures to test and evaluate the traceability and explainability of the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data. Conduct a data protection impact assessment.
- Anonymise and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).
- Implement procedures to test and evaluate the accuracy of the AI system, as well as the diversity and representativeness of the datasets and algorithm (post process model inference analysis, Human-in-the Loop decision flow). Regularly test the AI system for biases and take corrective actions if any are found.



6.2 Deanonimization – Geolocalisation of Images

Geolocalisation of images is the task of locating where an image was taken on earth when location metadata is missing or is incomplete. The task requires comparison of the target image with millions of images with location metadata to find the corresponding location. Automated tools are necessary for geolocalisation of images since humans perform poorly on this task.

Geolocalisation of images has improved considerably with recent AI techniques that identifies distinguishing features among huge amounts of images. Another trend is that the aerial perspective from publicly available satellite images is increasingly used to supplement ground level images. Recent developments combine the two perspectives for remarkable performance on a city scale.



Picture from Wikimedia, Creative Commons 4.0

Effectiveness (short-term perspective): High

Accuracy is highly dependent on the size of the geographic area. For city size areas, one kilometre precision is often possible with sufficient accuracy. Geolocalisation on earth is more difficult, especially of images with few features.

Robustness (short-term perspective): Medium

Robustness is highly affected by the trade-off between the geographic area size and accuracy. Huge variation in capture time and weather is also a problem. Information about scene type and context improves robustness.

Development (medium-term perspective): Active

Active commercial and academic research by many different actors. Recent efforts combine ground and satellite images. This improves performance considerably compared to only using ground images.

Projected future (long-term perspective): Promising

Image databases from social media and satellites will increase over time. Future development will increase performance and have high precision for even larger geographic areas.

TRL: Currently 4-6

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Surveillance Data
Publicly Available Information

Algorithms:
Natural Language Processing,
Deep Learning

**Risks:**

- To increase its accuracy, the reference database of the AI system needs to include a vast amount of non-personal data. Unnecessary personal data may inadvertently be processed, and this may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data.



6.3 Veracity Assessment – Disinformation Detection

Disinformation detection is the task of detecting fraudulent information that is intentionally spread to mislead people. Social media makes it easy to quickly spread large amounts of fraudulent information. The information may even be automatically generated in ways that are difficult for humans to detect. Automatic veracity assessment is necessary since manual assessment is costly and time-consuming.

Automatic veracity assessment consists of identifying claims that require assessment, finding sources that support or refute the claim, and assessing veracity using these sources. Recent efforts directly use source material to implicitly learn relevant features for all stages of veracity assessment. This improves performance considerably compared to previous approaches that only used contextual information (author, place of publication).

Discovery of the underlying intent may require an aggregated judgement from several detections of fraudulent information.



Picture from Wikimedia, Creative Commons 4.0

Effectiveness (short-term perspective): Medium

Effectiveness in detection of fraudulent information is probably highly context dependent. The effectiveness is high in simple contexts (reviews) and moderate in complex contexts (scientific facts).

Robustness (short-term perspective): Low

Most approaches only use one source (often Twitter). Robust detection of fraudulent information likely requires comparison of multiple sources.

Development (medium-term perspective): Very active

Active commercial and academic research by many different actors. Detection of fraudulent information is of interest for news agencies, public health, and businesses. Automatic selection of instances to label simplifies creation of datasets.

Projected future (long-term perspective): Hard to assess

Bots that automatically generate fraudulent information will make disinformation detection even more important. Probable to be an arms race between generation and detection of fraudulent information.

TRL: Currently 1-3

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Surveillance Data
Publicly Available Information

Algorithms:
Natural Language Processing,
Deep Learning

**Risks:**

- Depending on the envisaged use (e.g., the amount and category of subjected individuals, relevance of information scrutinized) the AI system may undermine the group anonymity and have chilling effects on the freedom of expression.
- If the accuracy is low, especially when retrieving evidence sources, the AI system may produce inaccurate outputs causing an unjustified interference with the freedom of expression.
- Depending on the language limitations in the datasets used or the envisaged use by LEAs, the AI system may systematically target individuals writing in certain languages and, thus, members of certain social groups. This may violate the right to non-discrimination.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Ensure the AI system is trained on diverse datasets, including various languages and cultural context. Implement procedures to test and evaluate the diversity and representativeness of the datasets and algorithm.
- Assess the algorithm's level of accuracy and implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop, etc); regularly test and update the AI system; in agreement with publishers, retrieve more evidence sources to avoid false positives or negatives.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).



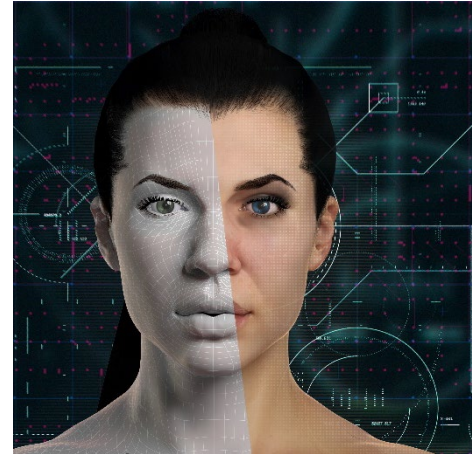
6.4 Detection of Synthetic Images

Today it is often impossible for a human to tell if an image has been computer generated. Therefore, we need tools to aid us in this task that automatically detect synthetic content in visual images.

Most detectors are probably only usable on synthetic images that are generated by a specific algorithm. A new detector is probably needed for each new generative algorithm. Detectors must be updated frequently due to the rapid development of generative algorithms. Detectors can be updated using either in-house expertise or by subscription to such a service.

Alternative countermeasures to synthetic images could include strong authentication techniques, which probably provide sufficient protection (but only for some cases).

Although detection of synthetic images has its limitations, the detectors will possibly succeed against less sophisticated actors who rely on out-of-the-box models (pre-trained and downloadable or available as a service) and are not able to modify them on their own.



Shutterstock, used with License

Effectiveness (short-term perspective): Medium

Synthetic images can be detected if they have been generated by algorithms that the detector is trained on. Some detectors can detect synthetic images that are generated with unknown algorithms (out-of-distribution images).

Robustness (short-term perspective): Low

Simple perturbations (cropping, compression, noise) reduce the likelihood of detecting synthetic images. There is currently no countermeasures to such perturbations. Synthetic images can also be tailored to avoid detection by known detectors.

Development (medium-term perspective): Very active

Active commercial and academic research by many different actors. Detection of synthetic images are of interest for news agencies and providers of images/photos. Likely to improve within a few years.

Projected future (long-term perspective): Hard to assess

Probable to be an arms race between generation and detection of synthetic images. AI-based tools are likely the only viable option for automatic detection of synthetic images.

TRL: Currently 4-6

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Publicly Available Information

Algorithms:
Natural Language Processing,
Deep Learning

**Risks:**

- In case of false negatives due to the low accuracy of the AI system, the presumption of innocence may be undermined whenever the outcome is considered undoubtedly incriminating.
- To increase its accuracy, the AI system needs to be trained with a vast amount of data, including personal and biometric data. This may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.
- Depending on the envisaged use (e.g., the amount and category of subjected individuals, relevance of information scrutinized) the AI system may undermine the group anonymity and have chilling effects on the freedom of expression.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Collaborate with the developer to improve the accuracy of the AI system (implement statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human in-the-Loop); assess the algorithm's level of accuracy and implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop); regularly test and update the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data. Conduct a data protection impact assessment.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).



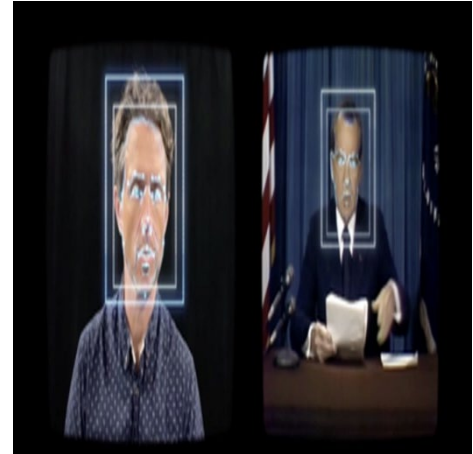
6.5 Detection of Synthetic Video

Today it could be impossible for a human to tell if a video has been computer generated. Therefore, we need tools to aid us in this task that automatically detect synthetic content in video.

A detector may use hand-crafted features, data-driven features, unique "fingerprints" of a generative algorithm, or artefacts in eye blinking, lip synching, facial landmarks, vocabulary, combinations of word classes or sound frequencies of speech. The features and artefacts are small enough that a human will not necessarily detect them.

Detectors must be updated frequently due to the rapid development of generative algorithms. Detectors can be updated using either in-house expertise or by subscription to such a service.

Alternative countermeasures to synthetic videos could include strong authentication techniques, which probably provide sufficient protection (but only for some cases).



Picture from Wikimedia, Creative Commons 4.0

Effectiveness (short-term perspective): Low

Today it is often possible to detect synthetic video with reasonable performance for well-known generative algorithms. However, the detectors do not generalise well to synthetic videos that are generated with unknown algorithms.

Robustness (short-term perspective): Low

Simple perturbations (cropping, compression, noise) reduce the likelihood of detecting synthetic videos. There is currently no countermeasures to such perturbations. Changing the generative algorithm will often thwart detection.

Development (medium-term perspective): Very active

Active commercial and academic research by many different actors. Detection of synthetic videos are of interest for news agencies and providers of videos. Likely to improve within a few years.

Projected future (long-term perspective): Hard to assess

Probable to be an arms race between generation and detection of synthetic videos. AI-based tools are likely the only viable option for automatic detection of synthetic videos.

TRL: Currently 1-3

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Publicly Available Information

Algorithms:
Natural Language Processing,
Deep Learning

**Risks:**

- In case of false negatives due to the low accuracy of the AI system, the presumption of innocence may be undermined whenever the outcome is considered undoubtedly incriminating.
- To increase its accuracy, the AI system needs to be trained with a vast amount of data, including personal and biometric data. This may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.
- Depending on the envisaged use (e.g., the amount and category of subjected individuals, relevance of information scrutinized) the AI system may undermine the group anonymity and have chilling effects on the freedom of expression.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Collaborate with the developer to improve the accuracy of the AI system (implement statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop); assess the algorithm's level of accuracy and implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop); regularly test and update the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data. Conduct a data protection impact assessment.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).



6.6 Language Models

Today it is often impossible for a human to tell if a text has been computer generated. Recent and ongoing leaps in automated text generation using AI language models (most notably GPT models) has sparked a debate and a demand for detectors of generated text.

Both generative and detection models are probability-based models. Generative models such as GPT choose the next word based on statistical assessments of the likelihood that it would appear next in the dataset the model is trained on. Detectors make similar statistical assessments based on a training set, but rather classifies the likelihood that a given text is written by a human or a language model. As such, the usefulness of classifiers is dependent on significant differences between human- and AI-written text. As generative models get increasingly better, these changes can be expected to decrease.



Generated using Bing Image Creator

The ongoing development in language models suggest that a focus on identifying AI-generated text is unlikely to be useful for LEAs. Rather, focus should be on how language models can be used to identify malicious language uses in cyber environments. This could be used for LEAs in both preventive and forensic purposes – such as identifying patterns in cyber scam conversations.

Effectiveness (short-term perspective): Medium-High

The already very good performance of language models, together with the rapid development of publicly available applications suggests language models can be of high use for LEAs in a short-term perspective. The main limiting factor is most likely availability of suitable training data.

Robustness (short-term perspective): Hard to assess

This depends heavily on application and availability of training data. With insufficient or bad quality data for the intended applications, there is a high risk for counterproductive tools.

Development (medium-term perspective): Very active

Active commercial and academic research by many different actors. Detection of synthetic text is of interest for education, national security agencies and other actors. Likely to improve continuously.

Projected future (long-term perspective): Hard to assess

Without implementation of new model architectures, the statistical models in use today are likely to reach a plateau in terms of quality and usefulness going forward.

However, where that plateau is, when it is reached and the consequences that arise along the way are difficult to speculate in as we are yet to see and understand the effects of tools such as ChatGPT on a large scale.

TRL: Currently 4-6

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Investigation

Data sources:
Surveillance data
Publicly Available Information

Algorithms:
k-Nearest Neighbour (k-NN)
Hidden Markov Model
GPT

**Risks:**

- To increase its accuracy, the AI system needs to be trained with a vast amount of data collected from various (online) sources, including social media posts and blog posts obtained via web scraping techniques. This may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.
- Since it undermines the group anonymity, the AI system may have chilling effects on the freedom of expression.
- Depending on the language limitations in the existing datasets or the envisaged use, the AI system may systematically target individuals writing in certain languages and, thus, members of certain social groups. This may violate the right to non-discrimination.
- Depending on the envisaged conditions of use, and especially is the accuracy is low, the use of the AI system can create unjustified interferences with the freedom of expression (e.g., if every (false) positive hit leads to the issuance of an order of content removal to the online platform).
- If the AI system continually learns, new patterns could be falsely flagged as anomalous until the system has sufficiently learnt.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. Ask AI developers to explain their algorithms' decision-making processes.
- Collaborate with the developer to improve the accuracy of the AI system (implement statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop); assess the algorithm's level of accuracy and implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop); regularly test and update the AI system.
- Implement procedures to test and evaluate the diversity and representativeness of the datasets and algorithm (post process model inference analysis, Human-in-the Loop decision flow). Regularly test the AI system for biases and take corrective actions if any are found.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).



6.7 Automatic Detection of Scammer Profiles

Due to the way scammer profiles are crafted and used, traditional preventive methods such as spam filtering are ineffective to cope with this new type of crime. This calls for new methods for detecting scammer profiles to prevent scams and aid victims. Increasing use of AI-generated materials (e.g. images, text for bios) for scammer profiles is expected to further increase the challenges associated with detection and make possible scams on a larger scale.

Online dating scams are, however, forced to uphold certain patterns in profile creation for the profiles to attract victims. This enables the use of AI-tools for identification/detection of potential scam profiles. Researchers have shown promising results with aggregated detectors, built on multiple specific classifiers for demographics, biographic text, and images. Each model can be trained on datasets of fake profiles to detect scammer profiles on dating sites. It is to be expected that the models for training will need continuous updating as scammers get access to more advanced generative techniques for producing scammer profiles.

Future uses for LEAs could include for example attribution of multiple scam profiles to one individual or criminal group or identifying different uses for scammer profiles.



Generated using Bing Image Creator

Effectiveness (short-term perspective): Low-Medium

The combination of multiple classifiers renders high accuracy results at best performance. However, the availability of datasets for training are scarce and hard to come by. As scammer profiles get better with generative AI, the major challenge is updating training data sets.

Robustness (short-term perspective): Low

As with most AI-detection systems, there are high risks of criminals developing evasion tactics. Adding more classifier types may add robustness against evasion. However, similar to the case with effectiveness, robustness of these models is reliant on access to sufficient amounts of updated training data.

Development (medium-term perspective): Very active

The development of generative AI-techniques, especially for text, voice, images and video is at the heart of AI-development at the moment. Lately big leaps have been taken in these areas and development is expected to continue.

Projected future (long-term perspective): Hard to assess

The future for automatic detection is tightly coupled with the quality of detectors for AI-generated content. However, for specific applications, such as scammer profiles, more easily identifiable patterns may suffice for automatic detection.

TRL: Currently 4-6

Additional information

General functionality:
Recognition
Prediction & Analysis

Application area:
Crime Prevention
Crime Investigation

Data sources:
Surveillance data
Publicly Available Information

Algorithms:
Naïve Bayes
Support Vector Machine



Risks:

- To increase its accuracy, the AI system needs to be trained with a vast amount of personal, often sensitive, data scraped by online social media (including dating websites). This may violate the right to data protection, especially the principles of lawfulness, data minimisation and the prohibition to profiling based on sensitive data.
- Since it undermines the group anonymity, the AI system may have chilling effects on the freedom of expression.
- Depending on the envisaged conditions of use, and especially if the accuracy and robustness is low, over-reliance on the AI system can create unjustified interferences with the freedom of expression and the right to access to a service (e.g., if false positive hits or incorrect recommendations lead to a ban from the online platform).
- Depending on the envisaged conditions of use, and especially when the detection of a scammer profile is followed by legal actions, the presumption of innocence may be undermined whenever the outcome is considered undoubtedly incriminating.
- If the AI system continually learns and analyses profiles, depending on the training data used, patterns in profile styles (e.g. using certain words, etc.) could result in increased numbers of false-positives until the system has sufficiently learnt about the patterns

Mitigation measures:

- Establish clear guidelines for the operation of the AI system, including mechanisms ensuring adequate and meaningful human oversight before any decision is taken. Ask AI developers to explain their algorithms' decision-making processes.
- Work together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); properly train and update the AI system and regularly test its reliability.
- Implement strict data protection regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).

Attention: The AI system shall not be used to flag individuals with an increased risk of becoming scammers, since this would be prohibited by the AI Act as a **predictive policing practice**.



6.8 Automatic Identification of Potential Scam Victims

Historically, choosing and mapping appropriate targets for spear phishing attacks or various forms of scams have been a laborious, manual task, including background research, identifying assets and weaknesses, as well as interaction with the victim to build rapport and trust. Today, antagonists can use various automated techniques (e.g., text and image analysis) based on information scraped from the internet to do large scale mapping and targeting of suitable victims. These analyses can then be aggregated into target maps to guide specific fraud attempts. Frequent users of e.g., social media will be more vulnerable for this type of automated mapping given the amount of data they produce about themselves on the internet.



Generated using Bing Image Creator

To work proactively and help protect likely targets from scammers, LEAs could potentially use similar AI-supported victim identification techniques combined with targeted information campaigns. Available studies show that individual characteristics (e.g., psychological traits and socio-demographics) can predict susceptibility for different kinds of cyber-scams. This knowledge together with automated identification could help potential victims in being cognizant of scam traits they would normally be more susceptible to. Combining classifications of potential victims and potential scammer profiles can also aid identification of high-risk matchings.

Effectiveness (short-term perspective): Low-Medium

On a group level, characteristics with predictive power seem to be fairly easy to identify, such as age, socioeconomic status etc. Based on these factors, identification of potential victims should be fairly accurate with available technology. However, the predictive power of these factors is unclear.

Robustness (short-term perspective): Medium-high

Available commercial techniques for mapping individual traits can be very accurate today. They are primarily used for targeted advertising, but also individualized education for example. It should be possible to adapt usage for targeted information campaigns regarding cyber scams within a foreseeable future.

Development (medium-term perspective): Very active

Big data analytics and behaviour mapping of individuals is at the heart of big tech businesses today. The development of such mapping techniques is constant. The availability for LEAs is unclear both in terms of technology access, and the availability for usage in accordance with data privacy laws and regulations.

Projected future (long-term perspective): Hard to assess

Can be expected to increase in use substantially as such tools and techniques become increasingly available and easy to use.

TRL: Currently 4-6

Additional information

General functionality:
Prediction & Analysis

Application area:
Crime Prevention

Data sources:
Surveillance data
Publicly Available Information

Algorithms:
Naïve Bayes
Support Vector Machine

**Risks:**

- The AI system needs to process a vast amount of personal, often sensitive, data scraped by online social media. This may violate the right to data protection, especially the principles of lawfulness, data minimisation and the prohibition to profiling based on sensitive data.
- Since it undermines the group anonymity, the AI system may have chilling effects on the freedom of expression.
- Depending on the envisaged conditions of use, and especially if the social media users cannot opt out from the profiling, the AI system can create unjustified interferences with the freedom of expression and the right to access to a service, as well as with the right to non-discrimination if certain social groups are particularly targeted.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. Ask AI developers to explain their algorithms' decision-making processes.
- Working together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the Loop). Regularly test and update the AI system.
- Implement strict data privacy regulations to protect group anonymity and individual privacy, as well as to limit the use of the AI system to targeted cases. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).

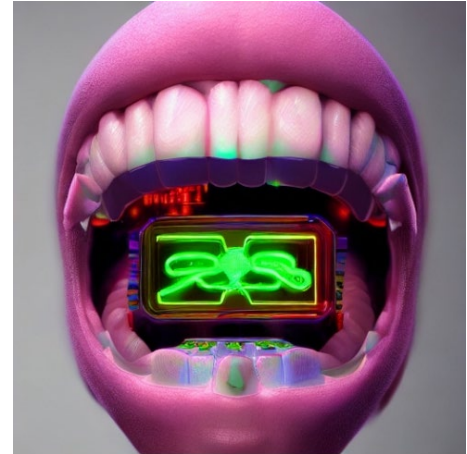
Attention: If the AI system is used to assess the risk of a natural person becoming a victim of a criminal offence, it will be qualified as **high-risk according to the AI Act**.



6.9 Detection of Voice Clones

Audio recordings can be deep faked to create audio content, such as “voice clones”. For deep fake audio. Machine learning engines are trained on different sounds to copy the voice patterns of some target individuals and then generate new audio clips with the same voice characteristics. Algorithmically generated voices alone can sound incredibly real and are indistinguishable to human ears and widely applied to produce realistic and natural deep fakes, exhibiting real threats. Voice clones can be used by scammers to contribute to a personal and trustworthy identity.

AI-based tools are likely the only viable option for automatic detection of synthetic voices. The tools can detect fake voices with reasonable performance for well-known generative algorithms. The detector is dependent on the learnt neuron behaviours and therefore not as effective with sounds that are generated with unknown algorithms. The AI approach can be a valuable solution to overcome several challenges by dealing with unlabelled data to work effectively in detection tasks. The technology is efficient and scalable.



AI generated image, <https://deepai.org/>

Effectiveness (short-term perspective): Low-Medium

Deep sonar detectors are dependent on the learnt neuron behaviours and therefore not as effective with sounds that are generated with unknown algorithms. The SSL-technology is efficient and scalable in solving the issues of supervised algorithms, but the detection rate is still low.

Robustness (short-term perspective): Low

Effective and robust detectors for synthesized fake voices are still in their infancy and are not ready to fully tackle this emerging threat. And the available versions are mainly trained to detect fake voice speaking English.

Development (medium-term perspective): Very active

Active commercial and academic research by many different actors. Detection of fake voices are of interest for news agencies, politicians, etc. The technique is likely to improve.

Projected future (long-term perspective): Hard to assess

As the threat of fake voices become a more triggered question due to its potential harm on political elections, fraud etc. it is likely that there will be an arms race between generation and detection of synthetic videos. AI-based tools are likely the only viable option for automatic detection of synthetic voices.

TRL: Currently 4-6

Additional information

General functionality:
Recognition
Prediction & Analytics

Application area:
Crime Prevention
Crime Investigation

Data sources:
Publicly Available Information

Algorithms:
Natural Language Processing
Unsupervised Learning
Deep Learning

**Risks:**

- In case of false negatives due to the low accuracy of the AI system, the presumption of innocence may be undermined whenever the outcome is considered undoubtedly incriminating.
- To increase its accuracy, the AI system needs to be trained with a vast amount of data, including personal data, often sensitive. This may violate the right to data protection and, especially, the principle of data minimisation.

Mitigation measures:

- Establish a robust legal framework that clearly defines the rights and responsibilities of all parties involved in the operation of the AI system
- Establish clear guidelines for the operation of the AI system. Ask AI developers to explain their algorithms' decision-making processes.
- Working together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop). Regularly test and update the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data. Conduct a data protection impact assessment.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed (e.g., usernames, email addresses, signatures in emails, IP addresses).



6.10 Detection of Crypto Currency Laundering

Criminals use crypto currency laundering to hide the illicit origin of funds, using a variety of methods. The most simplified form of bitcoin money laundering leans hard on the fact that transactions made in cryptocurrencies are pseudonymous. The same concepts that apply to money laundering using cash, apply to money laundering using cryptocurrencies.

Anti-money laundering (AML) refers to the laws, regulations, and procedures which guide financial institutions to prevent, detect and report money laundering activities. AI, machine learning (ML) and more specific natural language processing (NLP) can play a vital role in data management and analytics activities. The AML detection models are trained on non-reported alerts (i.e. transactions that are investigated but not reported), normal (un-investigated) transactions, and open information from multiple social networks to detect patterns of anomalous behaviours. The techniques are relevant for LEA by providing additional references (patterns of potential crypto currency laundering) with which a human investigator can make a final decision.



AI generated image, <https://deepai.org/>

Effectiveness (short-term perspective): High

AML detection accuracy is highly dependent on the size of the available data. The algorithms are for example able to detect abnormality where it identifies data points, events, and observations that deviate from a data set's normal behaviour. A NLP system can reduce the time and cost of operations by approximately 30% compared to previous manual approaches toward AML investigation.

Robustness (short-term perspective): Medium

Accessing AML data sets is an existing unsolved problem for the AML research community. There are a limited number of annotated money laundering datasets publicly available; this is a major problem that holds back AML research and especially for deep learning approaches.

Development (medium-term perspective): Very active

Anomaly detection can be performed for a variety of reasons and are steady developing. This improves performance considerably compared to previous approaches. By using AI and ML to detect suspicious transactions, given that a vast amount of data is available, the techniques can prove very useful.

Projected future (long-term perspective): Hard to assess

AI techniques are recent advances that could become standard approaches for AML. The system can boost the efficiency of money laundering detection without significant new capital investment. Depending on the data types, further anonymization may be needed to prevent the disclosure of clients' identities and other information.

TRL: Currently 4-6

Additional information

General functionality:
Recognition
Prediction & Analytics

Application area:
Crime Prevention
Crime Investigation

Data sources:
Publicly Available Information

Algorithms:
Machine Learning
Natural Language Processing
Deep Learning

**Risks:**

- To increase its accuracy, the AI system needs to be trained with a vast amount of (historical) data collected from various (online) sources, including social media posts. This may violate the right to data protection and, especially, the principle of lawfulness, data minimisation and the prohibition to profiling.
- To increase its accuracy, the AI system needs to be trained with a vast amount of historical data. This may violate the right to non-discrimination, if the data processed are not free of biases.
- Since it undermines the assumed internet anonymity (often searched while conducting blockchain-based transactions), the AI system may have chilling effects on the freedom of expression and the right to privacy.
- If the accuracy is low and, especially when AI users do not have access to information coming from other financial services, over-reliance on the AI system can unjustifiably impact individuals owning crypto currencies.
- If the AI system continually learns, new patterns in transactions could be falsely flagged as anomalous until the system has sufficiently learnt.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system, including mechanisms ensuring adequate and meaningful human oversight before any decision is taken. Ask AI developers to explain their algorithms' decision-making processes.
- Work together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); properly train and update the AI system and regularly test its reliability.
- Implement strict guidelines to protect group anonymity and individual privacy. Ensure that the AI system collects only the minimum amount of data necessary for its operation and implement mechanisms to delete the unnecessarily collected data. Conduct a data protection impact assessment.
- Implement procedures to test and evaluate the diversity and representativeness of the datasets and algorithm (post process model inference analysis, Human-in-the Loop decision flow). Regularly test the AI system for biases and take corrective actions if any are found.
- Ensure the auditability of the AI system to verify legal and technical compliance.



6.11 Using Drones for Handling an IED Incident

Unmanned Aerial Vehicles / Systems (UAV/S) have been tested and used as an operational source for EU LEAs for about a decade. The technique is increasingly being adopted for policing in many countries, as an aid for police officers to detect threats and respond to incidents with timely and cost-effective measures.

The identification, detection and classification of a suspected improvised explosive device (IED) can involve using many techniques, from sniffer dogs to X-ray scanners. Explosive substances present a serious risk to security, infrastructure and public safety so a broad approach using several existing detection methods is needed.

The European Defence Agency has developed AI-supported UAS and ground unmanned systems to autonomously search and detect IEDs – both in rural and urban areas. They have shown that AI can be used effectively in several aspects such as self-navigation, mission planning, team working and finally, detecting explosive devices. These tasks have been performed autonomously with minimal human intervention. Several sensors are being used with promising results. For example, neural networks were trained for each type of sensor signal processing, based on labelled data sets acquired during the project.



AI generated image, <https://deepai.org/>

Effectiveness (short-term perspective): Medium

The effectiveness of procedures to detect IEDs are dependent on the detection materials, environment and equipment, so new techniques are continuously explored to increase detection speed, precision and sensitivity.

Robustness (short-term perspective): Low

AI methods for explosive materials detection are not yet widespread but are undergoing rapid development. Effective and robust AI-supported IED detectors are still in their infancy and are not ready to fully tackle the current & emerging threat. AI-supported explosive material detection techniques require significant training data, which can be challenging to obtain due to the risks and safety concerns in handling and storing explosives.

Development (medium-term perspective): Active

Developments in technology, interdisciplinary collaboration, and the integration of AI techniques offer substantial opportunities for improving detection accuracy, decreasing false positives, and ensuring safer environments for individuals and society.

Projected future (long-term perspective): Promising

The identification and classification of IEDs is tightly coupled with the quality of training data for the detectors.

A European follow-up project has been selected from the European Defence Fund 2022 call and will take further the efforts on unmanned ground and aerial systems for IED neutralisation.

TRL: Currently 3-4

Additional information

General functionality:
Recognition

Application area:
Crime Investigation
Administration of Justice

Data sources:
Surveillance Data
Publicly Available Information
Previous Crimes

Algorithms:
Neural Network
Machine Learning

**Risks:**

- In case of false positives due to the low accuracy of the AI system, the right to privacy, the right to property and the presumption of innocence may be undermined whenever investigations and inspections are conducted.
- Depending on its functioning, the AI system may not provide enough information enabling the user to take informed decisions on follow-up actions.
- Depending on the envisaged conditions of use, and especially in case of untargeted use and when coupled with cameras, the AI system may have chilling effects on the freedom of expression and undermine the right to privacy.
- During the detection phase, the AI system may collect and process unnecessary personal data, thus violating the principles of lawfulness and data minimisation.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. Ask AI developers to explain their algorithms' decision-making processes.
- Working together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate explainability and human oversight measures to prevent over reliance and allow decision-making (Human-in-the-Loop). Regularly test and update the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of personal data necessary for its operation and implement mechanisms to delete the unnecessarily collected personal data.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed.



6.12 Facial Recognition

UAV/S have been tested and used as an operational source for European LEAs for about a decade. The technique is increasingly adopted for policing in many countries, as they can aid police officers to detect threats and respond to incidents with timely and low-cost services. By using a face recognition-based system LEA can identify criminals and use the technique for surveillance. A facial recognition system is a technology which involves the understanding of how the faces are recognized and detected, usually employed to match users through Identity verification services and computing facial features from a given image. The match is found by contrasting the individual face with the faces stored in the data base.



AI generated image, <https://deepai.org/>

UAV/S supported with sensors for face recognition can be built using various machine learning and deep learning algorithms, but each algorithm possesses varying advantages and disadvantages. Research has shown how hybrid methods can improve the performance of face recognition algorithms. UAV/S can utilize face recognition technologies to identify criminals and lost persons in crowds for a variety of reasons, for example remote surveillance during events where a large crowd is expected. Their use by P&LEAs is regulated in the new AI Act.

Effectiveness (short-term perspective): High

Hybrid methods, using different algorithms, can be used to improve the performance of face recognition algorithms. The comparative analysis of the algorithms facilitates the technologist to formulate algorithms with the utmost accuracy customized to the demands of the application. When the drone's camera angle is within 37 degrees, the accuracy is about 98.6%.

Robustness (short-term perspective): Medium

It is already difficult to identify unfamiliar people based on digital photographs of faces. This means that even if one could obtain high-resolution imagery of to-be-identified persons from drone-recorded footage, some errors could still be expected to occur.

Development (medium-term perspective): Active

Even though these algorithms have reached an accuracy of about 98%, researchers are still working on achieving full accuracy. Factors that affect identification performance from drone-recorded footage, is image quality, training data and additional person-related information from the body and gait.

Projected future (long-term perspective): Promising

Facial recognition is tightly coupled with the quality of training data for the detectors and the quality of drone-recorded footage. It is predicted that both will increase drastically in near time. It is anticipated that the use of UAV/S will increase over the next several years as technology develops, and new uses are found.

TRL: Currently 4,6

Additional information

General functionality:
Recognition

Application area:
Crime Investigation
Cyber Operations

Data sources:
Surveillance Data
Publicly Available Information

Algorithms:
Computer Vision



Risks:

- In case of false positives, the presumption of innocence may be undermined whenever the outcome of the AI system is considered undoubtedly incriminating.
- Depending on the quality of the training datasets used, the AI system may be biased and systematically target members of certain social groups. This may violate the right to non-discrimination.
- Depending on the envisaged conditions of use, and especially in case of widespread and/or untargeted use, the AI system may have chilling effects on the freedom of expression and undermine the right to privacy.
- To increase its accuracy, the AI system needs to be trained with a vast amount of data, including sensitive and biometric data. This may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.

Mitigation measures:

- Implement strict guidelines to ensure that the AI system only identifies certain individuals (e.g., suspects or victims of crimes) and implement mechanisms to delete the unnecessarily collected personal data. Conduct a data protection impact assessment and a fundamental rights impact assessment.
- Working together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop). Regularly test and update the AI system.
- Implement procedures to test and evaluate the diversity and representativeness of the datasets and algorithm (post process model inference analysis, Human-in-the Loop decision flow). Regularly test the AI system for biases and take corrective actions if any are found.
- Implement strict data privacy regulations to protect group anonymity and individual privacy. Anonymize and aggregate the training data processed.

Attention: In principle, the AI system shall not be used to identify individuals in a publicly accessible space and in real-time, since this would be prohibited by the AI Act as a **real-time remote biometric identification system**. LEAs can use the AI system only for the targeted search of individuals, if a fundamental rights impact assessment has been completed and a judicial authorisation has been granted.



6.13 Drone Use for Object Detection

Search & Rescue (SAR) teams are continuously seeking to improve their operations by developing new techniques to quickly find a person who is lost. UAV/S can be employed in object detection applications of SAR missions in remote areas. UAV/S assist teams in areas that are difficult and time consuming to be patrolled. The systems can be equipped with high-resolution cameras, thermal sensors, and other advanced technologies that provide real-time coverage and enable them to capture detailed images and data from the air. This data can be used for a variety of purposes, such as detecting objects and people. For example, drones equipped with thermal imaging cameras can be used to detect heat signatures and identify individuals who may be lost or injured in a wilderness area and/or water.



Picture from Wikimedia, Creative Commons 4.0

Object detection for UAS in SAR-operations can be equipped with algorithm YOLO (“You Only Look Once” which recognizes and detects several objects in a picture. The class probabilities of the discovered photos are provided by the object identification process in YOLO, which is carried out as a regression problem. Convolutional neural networks (CNN) are used to recognize items instantly. The approach just needs one forward propagation through a neural network to identify objects. The YOLO algorithm works by dividing the image into N grids that takes part in the detection and localization of the object it contains.

Effectiveness (short-term perspective): High

Usage of UAV/S is currently limited by difficulties such as satellite communication and cost, but the advent of programmable drones has enabled engineers to implement various technologies in an unmanned aerial vehicle that can be utilised in numerous fields. The effectiveness of the YOLO Algorithm for object detection is significant, with high speed, high accuracy and good learning capabilities.

Robustness (short-term perspective): Medium

The algorithm should consider the complexity of object detection on UAV/S footages due to factors such as image instability, low resolution because of platform movement, flying at high altitudes, angles and variations in camera position and so on.

Development (medium-term perspective): Active

AI-supported UAS have significantly improved real-time object detection in SAR-operations, balancing challenging trade-offs between accuracy, speed, portability and ease of deployment.

Projected future (long-term perspective): Promising

With future developments, the system will be further enhanced by using both optical and thermal cameras for 24 hour a day detection and their use in more adverse conditions.

TRL: Currently 4,6

Additional information

General functionality:
Surveillance

Application area:
Crime Investigation
Migration, Asylum, Border Control

Data sources:
Surveillance Data
Publicly Available Information
Previous Crimes

Algorithms:
Automated Vehicles

**Risks:**

- Depending on the envisaged conditions of use, and especially in case of untargeted use and when coupled with cameras, the AI system may have chilling effects on the freedom of expression and undermine the right to privacy.
- To increase its accuracy, the AI system needs to be trained with a vast amount of non-personal data. Unnecessary personal data may inadvertently be processed, and this may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.
- During the detection phase, the AI system may collect and process unnecessary personal data, thus violating the principles of lawfulness and data minimisation.

Mitigation measures:

- Establish clear guidelines for the operation of the AI system. This includes disclosing how the system works, the data it uses, and the basis for its decisions.
- Working together with the AI system developers, establish accuracy standards for AI systems to minimize false positives (statistical accuracy, confusion matrix, ground truth comparison, etc); implement appropriate human oversight measures to prevent over reliance (Human-in-the-Loop). Regularly test and update the AI system.
- Implement strict guidelines to ensure that the AI system collects only the minimum amount of personal data necessary for its operation and implement mechanisms to delete the unnecessarily collected personal data.



6.14 Chatbots

Today, technology exists that could serve a useful role in appropriate circumstances to help P&LEAs assist victims of crime. A significant increase of the technical support was seen during COVID-19 and the consequent lockdowns. Some organizations feature live chat functionalities, facilitated solely by human intervention, allowing for direct communication with support personnel. The support relies in deploying tech solutions, such as Chatbots, that enable survivors to seek support straight away and within their immediate surroundings.

Chatbots might play a role in relieving P&LEAs by supporting victims/survivors in situations where direct access to help is limited. Chatbots mimic human conversation by generating responses based on user input. Widely employed across industries like finance and healthcare, their usage has surged, especially with the launch of OpenAI’s ChatGPT, Microsoft’s CoPilot, and Google’s Gemini. Although still few, a handful of chatbot solutions have emerged, offering crucial information essential for victims/survivors as a possible alternative to live chats. Chatbots could effectively provide aid to victims by directing them to relevant institutions for specialist assistance and as a result, enhance incident documentation and management by gathering better and earlier data to support criminal cases.



Picture from Wikimedia, Creative Commons 4.0

Effectiveness (short-term perspective): High

The already very good performance of chatbots, together with the rapid development of publicly available applications, means that their use can be of high effectiveness for P&LEAs over the short-term. The main limiting factor is most likely to be availability of suitable training data and chatbots being too “linear” in their functionality to show emotional support.

Robustness (short-term perspective): Medium

This depends heavily on the application and availability of training data. With insufficient or bad quality data for the intended applications, there is a high risk of creating counterproductive tools. There is a clear resistance to fully automated chatbots. Instead, rule based chatbots are recommended. Chatbots are not recommended for conveying emotional support and providing full transparency.

Development (medium-term perspective): Active

Active commercial and academic research by many different actors is ongoing. Usage of chatbots to relieve national security agencies, and P&LEAs have been proposed and are likely to improve continuously.

Projected future (long-term perspective): Promising

Without implementation of new model architectures, the statistical models in use today are likely to reach a plateau in terms of their quality and usefulness. However, where that plateau will be, the time taken to reach it and the consequences that will arise along the way are difficult to speculate on as we are yet to see and understand the effects of these tools operating on a large scale and over time.

TRL: Currently 4,6

Additional information

General functionality:
Prediction & Analytics

Application area:
Crime Investigation

Data sources:
Surveillance Data
Publicly Available Information

Algorithms:
k-Nearest Neighbour (k-NN)
Hidden Markov Model
GPT

**Risks:**

- Especially when the host platform does not clearly indicate that the responder is a chatbot, the AI system may stimulate emotional attachment in the victim, thus undermining their freedom of thought, as well as their human agency and autonomy.
- Depending on the training data used and the algorithm's design choices, the AI system may generate an output that may be biased towards members of a certain social group or that further victimises the survivor. This may violate the right to non-discrimination and the right to human dignity.
- To increase the quality of its output, the AI system needs to be trained with a vast amount of non-personal data. Unnecessary personal data may inadvertently be processed, and this may violate the right to data protection and, especially, the principles of lawfulness and data minimisation.
- The AI system may collect and store personal data of victims (e.g., their name, their IP address and the conversation) by default. This violates the rights to privacy and data protection, especially the principle of lawfulness.

Mitigation measures:

- Ensure that users are informed when interacting with an AI system. Clearly indicate that the responder is an AI chatbot; provide disclaimers about the limitations of the AI system and its non-human nature.
- Implement mechanisms for human review and oversight. Human moderators can intervene when necessary to prevent harmful or inappropriate responses. Establish clear guidelines for human moderators to follow when reviewing AI-generated content.
- Implement procedures to test and evaluate the diversity and representativeness of the datasets and algorithm (post process model inference analysis, Human-in-the Loop decision flow). Regularly test the AI system for biases and take corrective actions if any are found. Inform users about the accuracy of the AI system.
- Implement strict data privacy regulations to protect individual privacy. Do not collect personal data (e.g., usernames, email addresses, signatures in emails, IP addresses) by default. Anonymize and aggregate the data processed.



7. Conclusions

This document provides recommendations for practitioners from P&LEAs, policymakers, researchers, and other relevant actors in the ecosystem surrounding AI for P&LEAs. Based on extensive work with European stakeholders via workshops and surveys, ALIGNER identified relevant scenarios of AI use by P&LEAs as well as potential misuse by criminals. The project identified capability enhancement needs, potential AI technologies to address these needs, as well as broader implications of the use and misuse of AI technologies, specifically in the ethical and legal domain, cybersecurity, and the potential future malicious use of AI.

Based on these findings, the project identified nine policy recommendations, together with the other projects of the EU AI cluster (popAI, STARLIGHT, AP4AI) as well as 19 research recommendations. These policy and research recommendations now need to be put into practice, for example, via hands-on implementation by the AI Office or via the inclusion of research recommendations in future EU research framework programmes.



8. References

- [1] L. Clutterbuck, "ALIGNER D2.2 Archetypical Scenarios and their Structure," H2020 ALIGNER, GA no. 101020574, 2022.
- [2] BSI, "How is AI changing the cyber threat landscape?," 2024. [Online]. Available: <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/How-is-AI-changing-cyber-threat-landscape.html>. [Accessed 21 09 2024].
- [3] BSI, "Generative AI Models: opportunities and Risks for Industry and Authorities," 2024. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4. [Accessed 21 09 2024].
- [4] ENISA, "AI Cybersecurity Challenges: Threat Landscape for Artificial Intelligence," 2020. [Online]. Available: <https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges>. [Accessed 21 09 2024].
- [5] M. Karresand and J. Reuben, "ALIGNER D3.4 Cybersecurity requirements structure for AI solutions," H2020 ALIGNER, GA no. 101020574, 2023.
- [6] M. Jarlsbo, N. Normelli and M. Svahn, "ALIGNER D3.3 Taxonomy of AI Supported Crime," H2020 ALIGNER, GA no. 101020574, 2024.
- [7] European Commission, "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," 21 April 2021. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>. [Accessed August 2024].
- [8] European Parliament and Council, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," 13 June 2024. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>. [Accessed August 2024].
- [9] M. Almada and N. Petit, "The EU AI Act: A Medley of Product Safety and Fundamental Rights?," Robert Schuman Centre for Advanced Studies, San Domenico di Fiesole, 2023.
- [10] European Union, "Glossary of summaries: Regulation," 2021a. [Online]. Available: <https://eur-lex.europa.eu/EN/legal-content/glossary/regulation.html>. [Accessed August 2024].
- [11] *Judgement of the Court of 10 October 1973. Fratelli Variola S.p.A. v Amministrazione Italiana delle Finanze (Case 34-73)*, 1973.



- [12] D. Casaburo and I. Marsh, "ALIGNER Fundamental Rights Impact Assessment," March 2023. [Online]. Available: <https://aligner-h2020.eu/fundamental-rights-impact-assessment-fria/>. [Accessed August 2024].
- [13] European Commission, "Artificial Intelligence - Questions and Answers," 1 August 2024a. [Online]. Available: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QAN_DA_21_1683_EN.pdf. [Accessed August 2024].
- [14] Ziouvelou et al., "popAI Policy Brief - 1st Year," Deliverable D1.6, H2020 popAI, GA no. 101022001, 2022.
- [15] Ziouvelou et al., "popAI Policy Brief - 2nd Year," H2020 popAI, GA no. 101022001, 2023.
- [16] L. Clutterbuck, R. Warnes and I. Marsh, "ALIGNER D2.3 Policy recommendations," H2020 ALIGNER, GA no. 101020574, 2022.
- [17] J. Laufs and H. Borrión, "Technological innovation in policing and crime prevention: Practitioner perspectives from London," *International Journal of Police Science & Management*, pp. 1-20, 2021.
- [18] D. Casaburo and I. Marsh, "ALIGNER D4.2 Methods and guidelines for ethical & law assessment," H2020 ALIGNER, GA no. 101020574, 2023.
- [19] European Parliament and Council of the European Union, "Directive 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purpose of the prevention, investigation, detection or prosecution of criminal offences," 4 May 2026. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680>. [Accessed 31 03 2023].
- [20] D. Pedreshi, S. Ruggieri and F. Turini, "Discrimination-aware data mining," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '08)*, New York, NY, USA, 2008.
- [21] C. Dwork, "Differential privacy: A survey of results," in *International conference on theory and applications of models of computation*, Berlin, Heidelberg, 2008.
- [22] Kairouz, Peter et al., "Advances and Open Problems in Federated Learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1-2, pp. 1-210, 2021.
- [23] UNSD, "Methodology - Standard country or area codes for statistical use (M49)," [Online]. Available: <https://unstats.un.org/unsd/methodology/m49/>. [Accessed 14 09 2022].
- [24] European COmmission, "AI Act | Shaping Europe's digital future," [Online]. Available: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>. [Accessed 24 09 2024].



- [25] S. Swain and L. Clutterbuck, "ALIGNER D2.1 Report on Law Enforcement Agency, Public, Scientific, Industrial and Ethical Stakeholder Involvement," H2020 ALIGNER, GA no. 101020574, 2024.



Annex A: Projects and Initiatives Mapping

Name	Brief Description	Website
AIDA - Artificial Intelligence and advanced Data Analytics for Law Enforcement Agencies	AIDA will develop a Big Data Analysis and Analytics framework equipped with a complete set of effective, efficient and automated data mining and analytics solutions to deal with standardised investigative workflows, extensive content acquisition, information extraction and fusion, knowledge management and enrichment through novel applications of Big Data processing, Machine Learning, AI and predictive and visual analytics. It will do so in a way that ensures societal benefits and consequences are integral part of design and deployment efforts.	https://www.project-aida.eu/
AP4AI – Accountability Principles for Artificial Intelligence in the Internal Security Domain	The AP4AI will create a global framework for AI accountability for policing, security and justice. This framework will be grounded in empirically verified accountability principles for AI as carefully researched and accessible standard, which supports internal security practitioners in implementing AI and Machine Learning tools in an accountable and transparent manner and in line with EU values and fundamental rights.	https://www.ap4ai.eu/
ARCSAR -Arctic and North Atlantic Security and Emergency Preparedness Network	Addresses the Arctic and North-Atlantic (ANA) region, preparing to cope with the security and safety threats that will result from increased commercial activity in the region including traffic through the northern passages, cruise traffic, and offshore oil and gas activity	https://arcsar.eu/
ARESIBO - Augmented Reality Enriched Situation awareness for Border security	The top priorities of ARESIBO will be scientific excellence and technological innovation. It will enhance the current state-of-the-art through technological breakthroughs in Mobile Augmented Reality and Wearables, Robust and Secure Telecommunications, Swarm Robotics and Planning of Context-Aware Autonomous Missions, and Artificial Intelligence (AI), in order to implement user-friendly tools for border and coast guards.	https://www.aresibo.eu/
CC-DRIVER - Understanding the drivers of cybercriminality, and new methods to prevent, investigate and mitigate cybercriminal behaviour	The CC-DRIVER project seeks to understand the drivers of cybercriminality and researches methods to prevent, investigate and mitigate cybercriminal behaviour.	https://www.ccdriver-h2020.com/project



CONNEXIONS - InterCONNECTed NEXt-Generation Immersive IoT Platform of Crime and Terrorism DetectiON, PredictiON, InvestigatiON, and PreventiON Services	CONNEXIONS aims to develop and demonstrate next-generation detection, prediction, prevention, and investigation services. These services will be based on multidimensional integration and correlation of heterogeneous multimodal data, and delivery of pertinent information to various stakeholders in an interactive manner tailored to their needs, through augmented and virtual reality environments.	https://www.connexions-project.eu/
CREST - Fighting Crime and TerroriSm with an IoT-enabled Autonomous Platform based on an Ecosystem of Advanced IntelligEnce, Operations, and InveStigation Technologies	CREST's overall objective is to improve the effectiveness and efficiency of LEAs intelligence, operation, and investigation capabilities, through the automated detection, identification, assessment, fusion, and correlation of evidence acquired from heterogeneous multimodal data streams	https://project-crest.eu/
CYCLOPES Fighting Cybercrime – Law Enforcement Practitioners' Network	CYCLOPES establishes a Europe-wide network to combat cybercrime.	https://cyclopes-project.eu
D4FLY - Detecting Document frauD and iDentity on the fly	The project focuses on enhancing the quality and efficiency of identity verification at border crossings in all modalities: land, air, and sea by providing faster and more secure border control solutions.	https://d4fly.eu/
DARENET Danube river region Resilience Exchange Network	DAREnet is building a dynamic multi-disciplinary community of practitioners, operating in a network of civil protection organisations. The network is supported by a broad range of stakeholders from policy, industry and research. Together they build an interdisciplinary ecosystem to foster synergies, innovation and its uptake across the Danube Region.	www.darenetproject.eu/
DARLENE - Deep AR Law Enforcement Ecosystem	Investigating how cutting-edge augmented reality (AR) technology can be deployed to help law enforcement agencies (LEAs) and first responders make more informed and rapid decisions especially in situations where time is of the essence. The project develops innovative augmented reality (AR) tools that aim to improve situational awareness when responding to criminal and terrorist activities	https://www.darleneproject.eu/
eNOTICE European Network of CBRNE Training Centres	The overall goal of the eNOTICE project is to establish a European network of CBRN training, testing and demonstration centres aiming at enhancing CBRN training capacity for improved preparedness and incident response through increased collaboration between CBRN training centres and practitioners' needs-driven CBRN innovation and research.	https://www.h2020-enotice.eu/



EU-HYBNET Empowering a Pan-European Network to Counter Hybrid Threats	The project is the 1st EU initiative which brings together pan-European practitioners and stakeholders to identify and analyse common challenges, and requirements to counter hybrid threats. It conducts research, highlights innovation initiatives, arranges training events to test innovations and makes recommendations for the uptake, industrialisation and standardisation of these innovations.	https://euhybnet.eu/
EXERTER Security of Explosives pan-European Specialists Network	EXERTER will provide practitioners with the operative knowledge and tools for enhancing the security of our society and to highlight innovative methods, tools and technologies, which can contribute in the fight against terrorism and serious crime. The aim is to help practitioners reach an improved capability, as well as to identify needs within standardisation and industrial development connected to Security of Explosives	www.exerter-h2020.eu
EXFILES - Extract Forensic Information for LEAs from Encrypted Smartphones	EXFILES will use software exploitation, hardware methods and combined methods to give law enforcement officials the tools and protocols for rapid and consistent data extraction in strict legal contexts.	https://exfiles.eu/
Fire-IN Fire and rescue Innovation Network	EU-wide one-stop shop for Fire-& Rescue Faster and cheaper access to the state-of-the-art Fire & Rescue technology for the whole of Europe	https://fire-in.eu/
FORMOBILE - From mobile phones to court – A complete FOREnsic investigation chain targeting MOBILE devices	Working in collaboration to create an end-to-end mobile forensic investigation chain, striving to improve digital safety, and security in the EU while respecting fundamental rights.	https://formobile-project.eu/
GRACE - Global Response Against Child Exploitation	GRACE aims to equip European law enforcement agencies with advanced analytical and investigative capabilities to respond to the spread of online child sexual exploitation material.	https://www.grace-fct.eu/
I-LEAD Innovation - Law Enforcement Agencies Dialogue	i-LEAD will build the capacity to monitor the security research and technology market to ensure a better matching and uptake of innovations by law enforcement agencies with the overarching aim to make it a sustainable Pan-European LEA network.	https://i-lead.eu/
ILEANET Innovation by Law Enforcement Agencies networking	ILEAnet aims to build a sustainable organisational Law Enforcement Agency (LEA) practitioners network focused on research & innovation addressing LEA challenges, together with a community of individuals interested to exchange and collaborate in this area. By encouraging such discussion between practitioners and experts from academia and industry, the project will stimulate LEA capabilities to influence, develop and take up research, development and innovation (RDI) that is useful and usable for LEAs, and thus help them to tackle the major challenges they face.	https://www.ileanet.eu/



iMARS - image Manipulation Attack Resolving Solutions	iMARS improves the operational capacity of passport application officers, border guards and forensic experts by providing Image Morphing and manipulation Attack Detection (MAD) solutions, Document Verification and Fraud Detection (DVFD) solutions, and by providing training, guidelines, share best practices and contribute to standardisation.	https://imars-project.eu/
INCLUDING Innovative Cluster for Radiological and Nuclear Emergencies	INCLUDING seeks to provide a full-fledged and comprehensive training in the RN security sector at European level. Starting from the existing training resources of the Partners in the Consortium, in most cases developed in the framework of EC projects, INCLUDING aims to enhance practical know-how and to boost a European sustainable training and development framework for practitioners in the RN Security sector.	https://including-cluster.eu/
INSPECTr - Intelligence Network and Secure Platform for Evidence Correlation and Transfer	The principal objective of INSPECTr will be to develop a shared intelligent platform and a novel process for gathering, analysing, prioritising and presenting key data to help in the prediction, detection and management of crime in support of multiple agencies at local, national and international level	https://inspectr-project.eu/
iProcurenetNet European Procurer Networking for security research services	iProcureNet aims to create an ecosystem of procurers, prescribers, legal advisors and other key stakeholders of security procurement, to share procurement trends and needs, and open pathways for joint procurement.	https://www.iprocurenet.eu/
LOCARD - Lawful evidence collecting and continuity platform development	automate the collection of digital evidence in any electronic format and medium. Its goal is to provide a comprehensive management approach to handle digital evidence to be presented in a court of law, alleviating many issues of current art and practice	https://locard.eu/
MEDEA Mediterranean practitioners' network	MEDEA is an EU funded Coordination and Support Action project the scope of which is to establish and further develop a regional Network of practitioners and other security related actors in the Mediterranean and the Black Sea region.	https://www.medeaproject.eu/
METICOS - A Platform for Monitoring and Prediction of Social Impact and Acceptability of Modern Border Control Technology	developing a platform that integrates information systems and networks of data sources in order to validate the efficiency and users acceptance of border control technologies. The proposed platform will provide metrics and KPIs to authorities and decision-makers, based on a number of independent variables: performance expectancy, effort expectancy, facilitating conditions, physical privacy, accuracy, information privacy, ethical and societal perceptions, securing positive societal impact and maximize border control process efficiency	https://meticos-project.eu/
NO-FEAR Network Of practitioners For Emergency medical systems and cRitical care	NO-FEAR will bring together a pan-European and beyond network of emergency medical care practitioners, suppliers, decision and policy makers to collaborate and exchange knowledge, good practices, and lessons learned.	http://no-fearproject.eu/



NOTIONES NetwOrk of an intelligence and security practitiOners with iNdustry and academia actorS	The NOTIONES project gathers actors from 15 European countries to develop European intelligence cooperation in the fight against crime.	https://cordis.europa.eu/project/id/101021853
PEN-CP Pan-European Network of Customs Practitioners	PEN-CP is 'a Novel Customs Innovation Boosting Network and On-line Platform to establish a customs practitioner network which facilitates translating customs security research and innovation ideas and requirements into scalable, viable solutions, technologies, and process improvements	https://www.pen-cp.net/
popAI - A European Positive Sum Approach towards AI tools in support of Law Enforcement and safeguarding privacy and fundamental rights	The core vision of pop AI is to foster trust in AI for the security domain via increased awareness, ongoing social engagement, consolidating distinct spheres of knowledge (including theoretical & empirical knowledge by academics & non-academics) and offering a unified European view across LEAs, and specialised knowledge outputs (recommendations, roadmaps, etc)	https://www.pop-ai.eu/
ROXANE - Real time network, text, and speaker analytics for combating organized crime	ROXANNE collaborates with Law Enforcement Agencies (LEAs), industry and researchers to develop new tools to speed up investigative processes and support LEA decision-making. The end-product will be an advanced technical platform which uses new tools to uncover and track organized criminal networks, underpinned by a strong legal framework.	https://www.roxanne-euproject.org/
STARLIGHT - Sustainable Autonomy and Resilience for LEAs using AI against High priority Threats	Law enforcement agencies' (LEAs) data-rich environments provide the opportunity to adopt artificial intelligence tools and capabilities that improve investigatory practices and limit the criminal misuse of AI. Through STARLIGHT, LEAs will collaboratively develop their autonomy and resilience in the use of AI for tackling major criminal threats.	https://starlight-h2020.eu/
SUSQRA - Protection against improvised explosive devices	SUSQRA aims at the development of an expert system to quantitatively assess the extent of damage caused by improvised explosive devices (IEDs) almost without using experiments.	https://www.emi.fraunhofer.de/en/business-units/security/research/susqra-sprengvorrichtungen-praevention-risikoanalyse.html
TAILOR - Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization	Purpose of building the capacity of providing the scientific foundations for Trustworthy AI in Europe by developing a network of research excellence centres leveraging and combining learning, optimization and reasoning.	https://tailor-network.eu/



Annex B: Additional information on the online surveys

To expand on the information gathered in ALIGNER’s workshops and obtain a more comprehensive picture of capability enhancement needs of P&LEAs as well as current trends in AI usage, ALIGNER conducted three online surveys over the course of the project. The following sections provide additional information on the first two surveys, held in 2022 and 2023. Information on the third survey, conducted in 2024, can be found in ALIGNER D3.3 [6].

First ALIGNER survey: 2022

The first ALIGNER survey was designed and conducted to gain an understanding of the capability enhancement needs perceived by those working in the field of law enforcement and policing. A further aim was to explore the potential challenges associated with integrating AI into law enforcement and policing further. The target group included practitioners working in this field as well as other professionals, e.g., from research institutions, who are concerned with the topics of AI, law enforcement, and policing (see Figure 12).

The survey consisted of a total of 25 questions, some of which were asked in a closed format with predefined answer options and some in an open format. This mixed approach was chosen to ensure an objective evaluation of the results on the one hand (closed questions) and to give participants the opportunity to address additional aspects on the other (open questions). Data-sensitive and personal questions, such as age or gender, were kept optional if this information was not crucial for gaining knowledge. All other questions were either provided with the option to skip the question or tick the “Not sure” option. This approach was chosen to counteract overload, e.g., in case of misunderstanding or not understanding the question, and to support higher data quality.

The survey was open from May 2022 on, and responses received by August 2022 were included in the roadmap. A three-months period was therefore set for the collection of survey responses. To gather opinions and experiences from the dedicated target group, a snowball sampling method was used. The survey was disseminated among ALIGNER’s advisory board members as well as related projects and their respective networks. Additionally, the link was published on LinkedIn.

It is important to note that the survey results only reflect the opinions of the sample studied and that no conclusions can be drawn for the entire population of interest. Furthermore, the identified capability enhancement needs in which AI could be of use are considered from a one-dimensional perspective that does not take into account all the potential consequences that would result from the application of AI in these areas. The initial collection of challenges in the survey scratches some important issues to consider and provides an impetus to discuss these within society as a whole.

Aim

- ❖ Understand the capability enhancement needs of law enforcement and policing

Target group

- ❖ Practitioners and professionals working in the field of law enforcement and policing

Scope

- ❖ 25 questions in closed and open format

Timeframe

- ❖ May - August 2022

Figure 12: Conditions of the first ALIGNER survey.



Demographic information on the survey sample

The survey was completed by a total of 53 respondents, of whom 16 (32%) were female and 34 (68%) were male³⁸ (Figure 13). The age distribution among the participants was quite balanced, with the largest part of the sample (35%) being between 45 and 54 years old and 20% representing respectively the age groups 25 to 34 years, 35 to 44 years, and 55 to 64 years. A small proportion of the sample (2%) was in the age groups 18 to 24 years and 65 years and older³⁹ (see Figure 13 for totals).

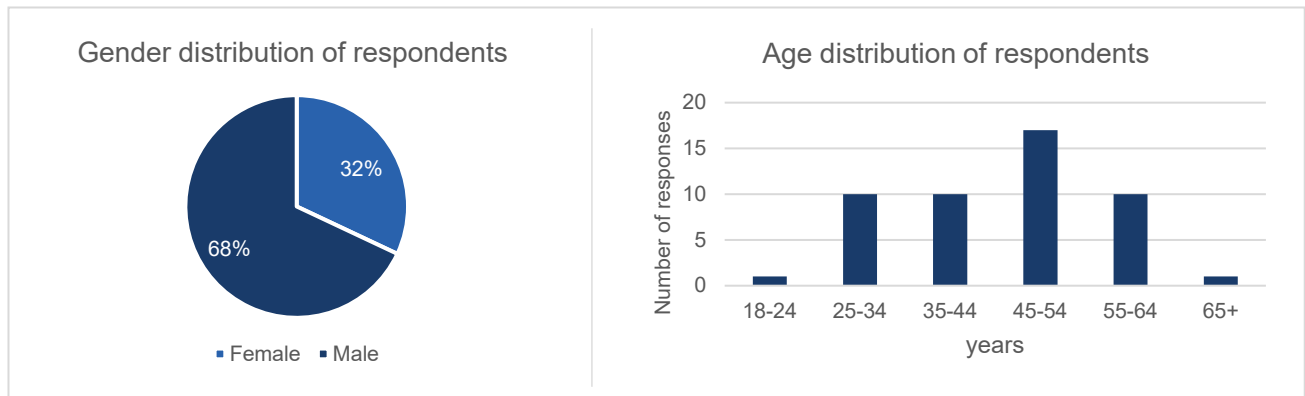


Figure 13: Results of the optional questions “What is your gender?” and “What is your age?”, first ALIGNER survey, 2022.

The sample consisted of 19 people working in law enforcement and policing (as practitioners) and 29 people working in research and academia. Two persons indicated “civil society” and “other” as their work organisation, and one person works in industry (Figure 14). The distribution of countries represented by participants’ work organisations is shown in Figure 15. Most participants (25 persons) were from Southern European countries (Greece, Italy, Kosovo, Portugal, Spain), followed by 14 persons working in Western European countries (Belgium, France, Germany, Netherlands). A proportion of 9 people work in Northern Europe (Estonia, Ireland, Lithuania, Sweden, UK) and 5 persons work in Eastern Europe (Bulgaria, Poland, Slovakia). This loose division into four geographical regions of Europe is based on a methodology of the Statistics Division of the United Nations Secretariat [22].

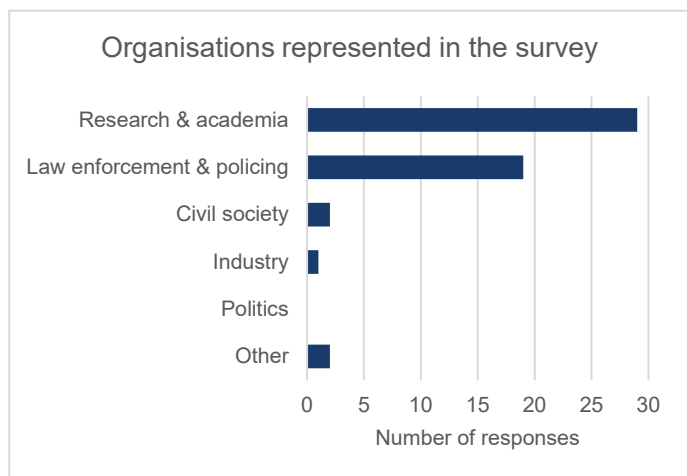


Figure 14: Results of the question “In which type of organization do you work?”, first ALIGNER survey, 2022.

³⁸ The question about gender was optional, so that n in this question deviates slightly from n total.

³⁹ The question about age was optional, so that n in this question deviates slightly from n total.

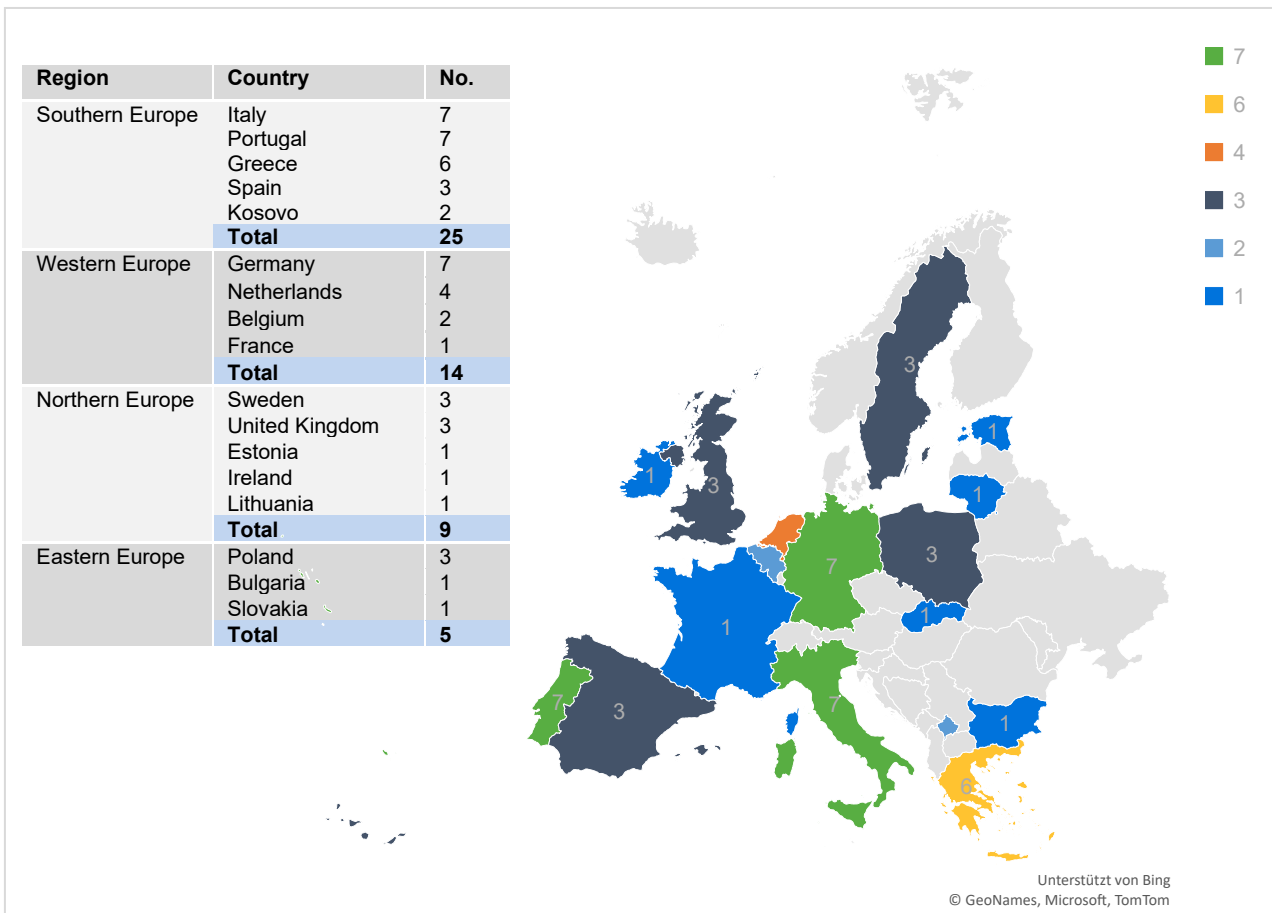


Figure 15: Results of the question “In which country is your organization based?”, first ALIGNER survey, 2022. *Note: Numbers on the map represent number of responses (e.g. “1” = 1 person working in country x, “2” = 2 persons working in country y).*

Unprioritized and categorized answers on potentials and challenges

Table 2 and Table 3 provide the raw data obtained from the survey on the questions “Where do you think AI can be applied immediately and would bring the greatest (immediate) benefit?” and “What do you think are the biggest challenges to introducing AI into law enforcement and policing?”. Answers not given in English have been translated, no other processing of the data has been performed.

Table 2: Original answers to the question “Where do you think AI can be applied immediately and would bring the greatest (immediate) benefit?”, first ALIGNER survey, 2022.

Where do you think AI can be applied immediately and would bring the greatest (immediate) benefit?

DNA analysis, object recognition, automated picture to picture comparison,
Digital forensics

1. Biometric recognition and identification, especially analyzing DNA traces (find similar traces in the data bank). Face recognition works quite well already. 2.Prevention of crimes within digital domains.

Detection and prevention of crimes and threats within the digital domain...

data handling and information handling processes

medicine

Data analytics across multiple police sources/systems



Data analytics, in particular filtering of relevant data

In intelligence and operational efforts

social media analysis and early detection of terrorism-related online crime (e.g. recruitment, propaganda, etc.)

digital forensics

Digital forensics

Data and information handling processes, Digital forensics

Computer Examinations and Phone examination.

Data Management

Cyber crimes

Face recognition (all capabilities of deep learning applied to images). Introducing advanced control tools (like deep reinforcement learning) to devices.

Monitoring of Social networks for detecting hate speech, radicalisation,...

evasion of excise duties

Risk assessment, social media analysis

information handling: making the best use of information that LE already has available (eg, criminal reports)

Digital forensics, passport and ID-related tasks

Fingerprints recognition and matching

Drones used for rescue operations

Detection and prevention of crimes and threats occurring outside the digital domain

Collecting and organization of data

Detection and prevention of crimes and threats within the digital domain

Biometric recognition and identification

Automation of search and data correlation procedures

video surveillance; usage of IoT devices (swarm optimisation)

Mass crime processing and analysis, facial recognition, pattern recognition

predictive policing, automated mapping of crimes and data analysis

AI and its tools can be applied immediately in identifying and predicting threats in cybersecurity processing a large amounts of data increasing both the speed and accuracy of decision-making processes.

Table 3: Original answers to the question "What do you think are the biggest challenges to introducing AI into law enforcement and policing?", first ALIGNER survey, 2022.

What do you think are the biggest challenges to introducing AI into law enforcement and policing?

general level of digitalization of the LEA, Trustworthiness

tendering process of public bodies, lack of transparency in the result creation process, IT legacy systems

The need for police & law enforcement agencies to understand what problems and challenges they face that technology can assist them with, followed by where and how they can obtain and operate the most suitable and ethically acceptable solutions

people are hesitant to use new technologies and need to have enough trust into the system, transparency about the AI system and how it was created to rule out implementing human biases into the system, to prevent harm by the system/ that the system gets hacked and used against you

Decision makers' lack of knowledge and concerns about AI technologies. Decision makers do not understand the technology.

Data protection.



Challenges are related to the psychology of the individual and groups to adopt disruptive technologies...

legislation and knowledge

Having a clear and precise definition of AI and communicating that definition to public & policy makers

Lack of labelled data for training of AI, experts for labelling are already a scarce resource

To be aware of the ethics and LAW

harmonisation of the regulation across countries; acceptance and training at operational level

privacy issues

Preventing algorithmic bias

data protection

Transparency and explainability (in a wide sense)

Internal safety rules, money, law gaps.

Capacity building. Capacity of understanding and using AI.

AI will only give "hints", the police person is the only one who can decide whether this is relevant or not...

Therefore according to me the largest challenges is to make AI useful for human (and gives them elements to improve their work, not replace their brains!)

To use it as the evidence for the court purposes

Operators to understand and accept benefits from AI

1. The human confidence of an AI system to be used in a legal issue. 2. How to legally manage a fail of an AI system that produces damages of any kind.

Trust

GDPR, bad reputation of AI

Crime detection, logistics

staying within legal & ethical boundaries; data governance (internal processes related to data quality, management, standardisation, etc)

Law and ethical principles, defining the exact ways that AI can and cannot be utilised

The biggest challenges are (1) making sure that the tools developed do not infringe privacy and lead to mass surveillance, (2) using AI into law enforcement and policing requires handling uncertainty (for instance in Computer vision) (3) Interpreting laws is subjective and is dependent on the situation, being able to handle this margin between right and wrong is a human trait that is difficult to enforce with an AI

Lack of transparency / human in control.

NLP

While AI can enhance capabilities as given above, this does not mean it is a good use of AI for society.

That it is not in breach of human rights or legislation

Legal framework

Privacy right compliance

the shift to new knowledge

Training of personnel and acquisition of tailored equipment to allow the usage of advanced AI capabilities

Using the AI technologies in a responsible way (e.g., fairly to every citizen).

rule of law - gdpr regulations and data protection issues

Proposals are often made by companies that miss the target of law enforcement or do not have much benefit (e.g. Precobs = making crime forecasts)

protection of privacy, chilling effect, human oversight

The biggest challenge facing the AI into law enforcement and policing is the need to reconcile AI's data with the with the human right to privacy taking into consideration current privacy legislation and culture.



Second ALIGNER survey: 2023

The second ALIGNER survey was designed and conducted to gain an understanding on how the recent emergence of LLM-powered AI technologies, like ChatGPT, had started to affect the work of P&LEAs. The target group again included practitioners working in this field as well as other professionals, e.g., from research institutions, who are concerned with the topics of AI, law enforcement, and policing (see X).

The survey consisted of a total of 19 questions, some of which were asked in a closed format with predefined answer options and some in an open format. As in the first survey, this mixed approach was chosen to ensure an objective evaluation of the results on the one hand (closed questions) and to give participants the opportunity to address additional aspects on the other (open questions). Data-sensitive and personal questions, such as age or gender, were again kept optional if this information was not crucial for gaining knowledge. All other questions were either provided with the option to skip the question or tick the “Not sure” option.

The survey was open from May 2023 on, and responses received by August 2023 were included in the roadmap. A three-months period was therefore set for the collection of survey responses. As for the first survey, a snowball sampling method was used, with the survey being disseminated among ALIGNER’s advisory board members as well as related projects and their respective networks. Additionally, the link was again published on LinkedIn.

Demographic information on the survey sample

The survey was completed by a total of 65 respondents, of whom 16 (25%) were female and 49 (75%) were male⁴⁰ (Figure 16). Regarding the age distribution, most respondents fell within the age ranges of 35-44 (29%) or 45-54 (29%), followed by 55-64 (18%), and 25-34(15%). As in the first survey, a small proportion of the sample fell within the age groups of 18-24 (5%) and 65 or older (3%)⁴¹ (see Figure 16 for totals).

Aim

- ❖ Understand how recently emerged AI technologies impacted the work of P&LEAs

Target group

- ❖ Practitioners and professionals working in the field of law enforcement and policing

Scope

- ❖ 19 questions in closed and open format

Timeframe

- ❖ May - August 2022

⁴⁰ The question about gender was optional, so that n in this question deviates slightly from n total.

⁴¹ The question about age was optional, so that n in this question deviates slightly from n total.

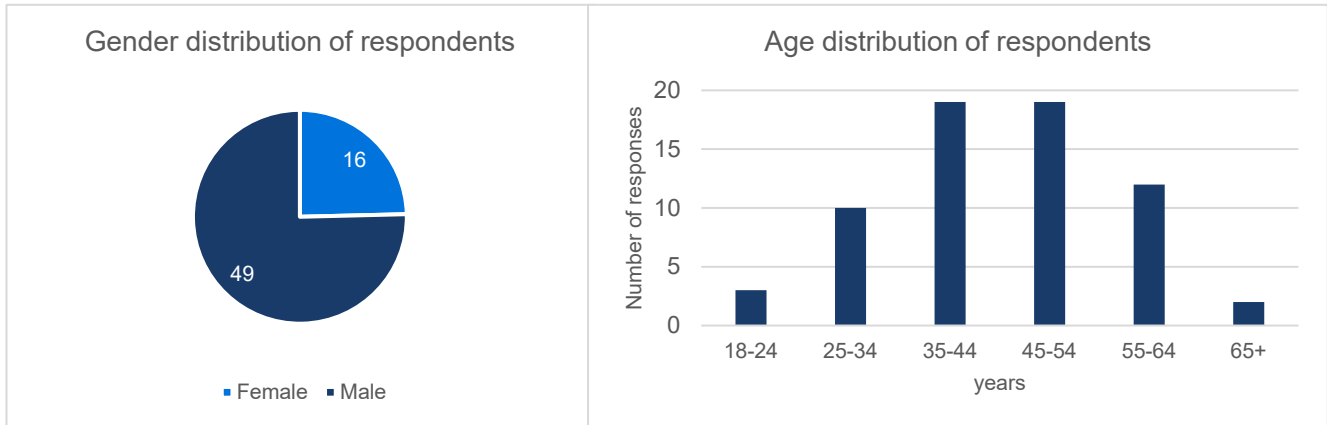


Figure 16: Results of the optional questions “What is your gender?” and “What is your age?”, Second ALIGNER survey, 2023.

The sample consisted of 30 people working in law enforcement and policing (as practitioners) and 24 people working in research and academia. Two persons indicated “civil society” while three indicated “other” as their work organisation, and seven persons work in industry (Figure 17). The distribution of countries represented by participants’ work organisations is shown in Figure 18. Most participants (22 persons) were from Southern European countries (Greece, Italy, Kosovo, Portugal, Serbia, Spain), followed by 18 persons working in Western European countries (Belgium, France, Germany, Netherlands, Switzerland). A proportion of 18 people work in Northern Europe (Ireland, Lithuania, Norway, Sweden, UK) and 5 persons work in Eastern Europe (Moldova, Poland, Romania). The division into four geographical regions of Europe is again based on a methodology of the Statistics Division of the United Nations Secretariat [22].

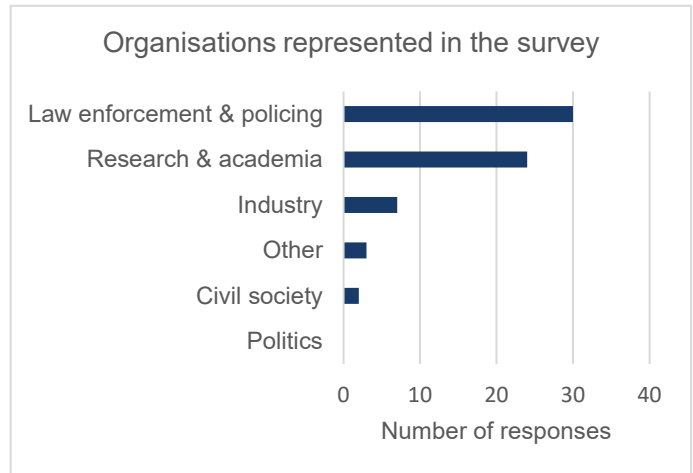


Figure 17: Results of the question “In which type of organization do you work?”, second ALIGNER survey, 2023.

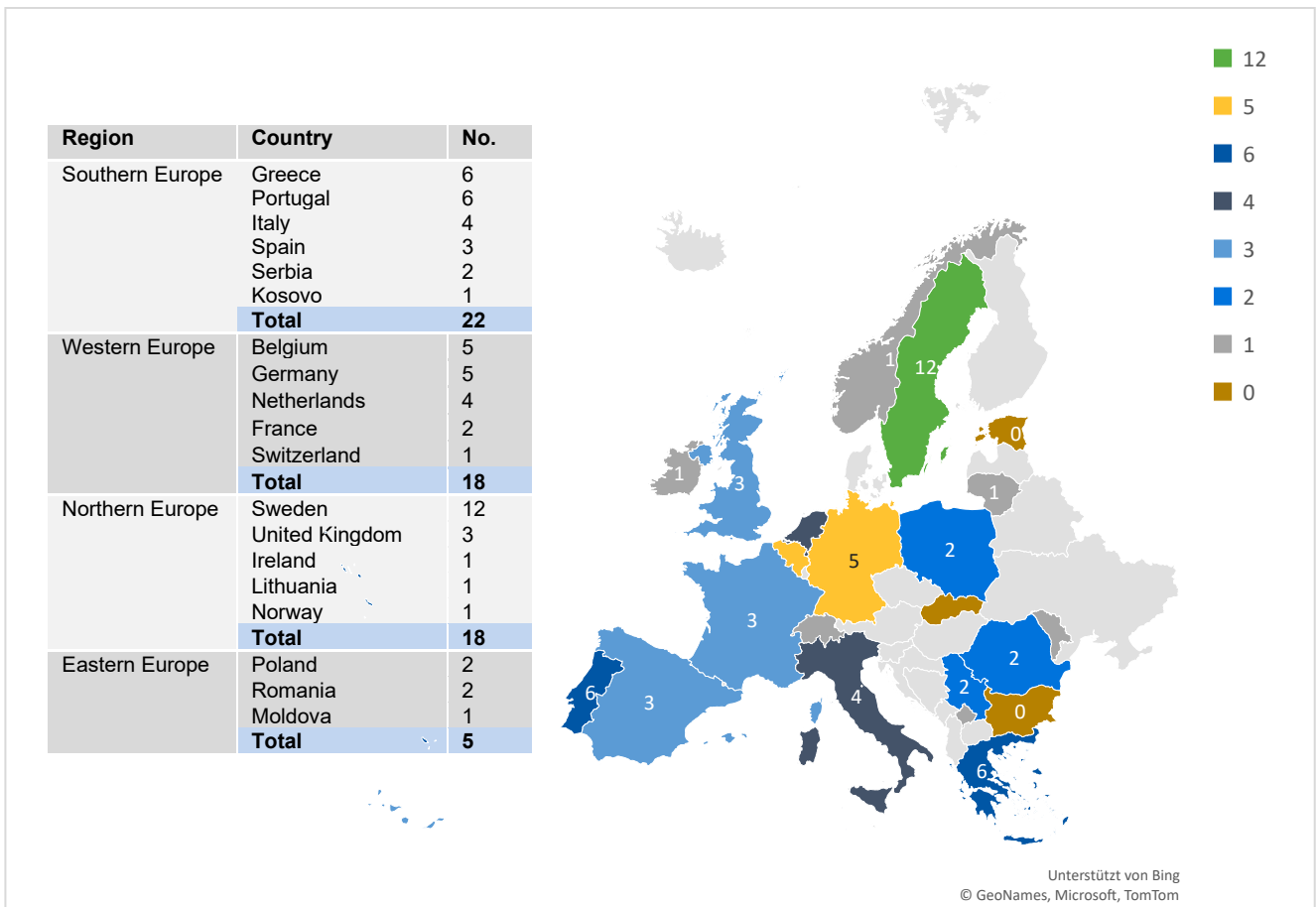


Figure 18: Results of the question “In which country is your organization based?”, second ALIGNER survey, 2023. *Note: Numbers on the map represent number of responses (e.g. “1” = 1 person working in country x, “2” = 2 persons working in country y). Countries with “0” had previously responded in the first survey.*

Unprioritized and categorized answers on examples for use of emerging AI technologies

X and Y Table 3 provide the raw data obtained from the survey on the questions “If you know any, please provide specific examples, or explain more generally how the recently emerging AI applications, tools, or technologies are or will be used by police and law enforcement?” and “If you know of it, can you provide examples of how criminal patterns have changed with the recent emergence of AI applications?”. Answers not given in English have been translated, no other processing of the data has been performed.

Table 4: Original answers to the question "If you know any, please provide specific examples, or explain more generally how the recently emerging AI applications, tools, or technologies are or will be used by police and law enforcement?", second ALIGNER survey, 2023.

If you know any, please provide specific examples, or explain more generally how the recently emerging AI applications, tools, or technologies are or will be used by police and law enforcement?
Extensions to ANPR like cell phone use detection, anomalous behaviour detection, illicit cargo detection, ... but also illegal border crossing detection
Border and surveillance/ data and information
Clearview AI
Big data tools, video analysis tools looking for unusual potentially risky situations...
Image/data evaluation, prediction, IT system analysis



Facial recognition, data, voice and video analysis

Border surveillance and bordercontrol

Early warning systems for extreme climate events

My company develops facial recognition systems which have been used by police and law enforcement for years with tendency to increased application.

Searching for information

Automation of some of the manual tasks like transcription, translation etc

Assist in system development, Translations, Employment Interview Questions

AI enhanced computer vision

Departments that have traditionally used more analogue methods in their investigative work are now used as test pilots for various efficiency experiments involving AI tools. The hope is to be able to demonstrate the benefits of AI in order to accelerate further AI transformation of the organisation as a whole.

Face Reid, Person Reid, Logo Detection and Search

In many cases, results from AI will expedite results that, otherwise, law enforcement would still get, but with more time

Language models (e.g., chatgpt)

From agencies I know, AI is being used for image processing (e.g., face recognition, projectile identification in ballistic), voice processing (e.g., speech to text)

Table 5: Original answers to the question "If you know of it, can you provide examples of how criminal patterns have changed with the recent emergence of AI applications?", second ALIGNER survey, 2023.

If you know of it, can you provide examples of how criminal patterns have changed with the recent emergence of AI applications?

Processing minors' normal images through an AI algorithm converted them to naked bodies.

Audio-Deepfakes, synthetic images

AI research and developed technologies could also facilitate criminal behaviours, shaping a new form of AI-Crime (e.g. Fake content generation for blackmailing and general harassment against individuals, weaponizing driverless vehicles, drones and other UAVs for illegal activities, etc., manipulating face recognition system, using social bots/sexbots for emulating sexual offences etc.)

Manipulation of the data analysis

Sophisticated fraud techniques, advanced threats and attacks on security systems, private data breaches, manipulation of sound, images and videos.

To use another identity

Criminals know about the existence of facial recognition systems and therefore hide their faces

Online fraud (semi automated)

The use of GPT models, deep fakes (images, audio & video) in several fraud cases.

Criminals have access to cheap and scalable solutions for content and communication to scam people

Computer generated AI enhanced CSAM

Precise localization using spatial data control, as well as data control and tracking within specific indoor environments. Real - time data control.

First in the cybercrime area, they started using AI tools to penetrate systems.