ALIGNER D2.4

Policy recommendations:

Second iteration, October 2023

Deliverable No.	D2.4
Work Package	WP2
Dissemination Level	PU
Author(s)	Lindsay Clutterbuck (CBRNE)
Co-Author(s)	-
Contributor(s)	Donatella Casaburo (KUL); Fredrik Bissmuss (FOI) Richard Warnes (CBRNE)
Due date	2023-09-31
Actual submission date	2023-10-15
Status	Final
Revision	1.0
Reviewed by (if applicable)	Michael Chin (LEAAB), Zvonimir Ivanovic (LEAAB), Clare Thornley (SIEAB), Shaban Buza (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 aim of Project ALIGNER is to bring together *"…European actors concerned with Artificial Intelligence (AI), Policing and Law Enforcement and to collectively identify and discuss needs for paving the way for a more secure Europe in which Artificial Intelligence supports LEAs while simultaneously empowering, benefiting and protecting the public."*

ALIGNER has been set a number of Specific Objectives (SO) relating specifically to policy recommendations. They key one sets out the requirement for them to be *"…tailored to the operational needs of the law enforcement sector…supported by [identified] capability gaps…[and] provide an overview of them from a societal and ethical perspective."*[1] This document is the second of three deliverables concerning policy recommendations. The first one, D2.3, was submitted in September 2022 and the third and final one, D2.5, will be submitted at the end of the project in September 2024.

This second document builds on the original six ALIGNER policy recommendations in a number of ways and commences by reviewing relevant events during the twelve-month period that has elapsed since then. AI technology has demonstrated its potential to make fundamental advances almost overnight and consequently, to increase the complexities that policy makers and police and law enforcement agencies (P&LEAs) alike must face in order to respond to it. In addition, the rapid advancement of AI capabilities and its open accessibility has driven the topic of AI technology onto the agendas of mainstream politicians and into the awareness of the media and the general public.

During this same period, a fundamental EU legislative building block in the response to the development of AI technology, namely the EU Artificial Intelligence Act, has slowly moved forward. The EU single, original AI Act Proposal forward by the EC in April 2021 currently consists of, in effect, three different proposals; the original proposal, the 'general approach' to it adopted by the EU Council in November 2022 and the 'negotiating position' proposed by the EU Parliament in July 2023. For the AI Act to become EU law all three bodies must negotiate to agree on its final wording and to date, this stage has not been reached.

The overall effect of both these drivers of change has been to make the environment within which P&LEAs operate legislatively uncertain while at the same time, the potential scale and scope of the AI technology challenges they already face have rapidly expanded.

ALIGNER forms part of a cluster of three projects relating to AI technology in the context of policing and law enforcement.[2] Having derived the ALIGNER policy recommendations in 2022, the next step in their development has been to examine whether or not they were in concordance with the other projects. To achieve this, the six initial ALIGNER policy recommendations were subjected to a cross-project comparison to examine their congruency with a number of relevant recommendations selected from those proposed by the other projects during the latter part of 2023.

In summary, twenty-eight selected policy recommendations were assessed and were found to be congruent with the ALIGNER policy recommendations a total of forty-two times. This total includes

[1] ALIGNER Description of Work, Part B-4 and B-5

[2] The other two AI cluster projects are 'PopAI' (*European Positive Approach towards Artificial Intelligence tools in support of Law Enforcement & Safeguarding Privacy & Fundamental Human Rights*) and 'STARLIGHT' (*Enhancing the EU's strategic autonomy in field of Artificial Intelligence for LEAs).* The third addition to the comparison used the recommendations and research needs identified by ENISA (European Union Agency for Cybersecurity).

instances where the same recommendation was judged to be applicable to two, or even three, ALIGNER policy recommendations. The distribution of the congruent recommendations across the three projects is not equal, with twenty-four found within Project PopAI, ten in Project STARLIGHT and eight within the recommendations from ENISA.

On examination from a high level cross-project perspective, the congruency assessment yielded a number of insights into the overall policy recommendation environment of them all, including those of ALIGNER. In essence, these insights are predominantly associated with the need to provide certainty for P&LEAs relating to how they can acquire and utilise trusted AI technology in a holistic way, and how they can do so within a regulatory environment that both supports and facilitates P&LEA operations and functions. A second, smaller category of suggestions refers to specific requirements to enhance the functions and activities of P&LEAs.

The most important policy recommendations were aggregated and then summarised as follows:

- ❖ The procurement, utilisation, in-service development and use of AI technology tools by P&LEAs or on their behalf should be approached in a holistic way. This should be based on a common European understanding of AI technology and carried by P&LEAs utilising common and harmonised EU legal directions and guidelines

- ❖ For P&LEAs to effectively engage with AI technology and its potential capabilities, there is a need for an EU-wide framework to enable, enhance and institutionalise how they do so. It should be designed for this purpose and utilise elements of primary legislation, statutory regulations, guidelines and protocols, systems and structures, processes and procedures

- ❖ The role of ethics in the P&LEA approach to AI technology and tools is seen to be of fundamental importance and should be governed by a clear and consistent 'ethical framework' of principles and practices

- ❖ AI technology tools for P&LEAs in the EU should be legally compatible and functionally interoperable. To achieve this, they should be evaluated (for legality, ethical compliance and functional performance). Consequently, there is a need for agreed standards to enable them to be assessed and compared

- ❖ The establishment by the EU of a single body or organisation to act as a nexus between rapidly developing AI technologies and their impacts on P&LEA functions and responsibilities could bring many benefits. Two of its suggested functions are as an Observatory to monitor AI and cybersecurity threats (ENISA) and as a central AI Systems Registry of each one used by P&LEAS across the EU (STARLIGHT)

  Other functions could also be envisaged for it, for example as a source of impartial and accurate advice and guidance for P&LEAs on issues relating to AI technology and its use. This type of approach would contribute to a common European understanding and enhance P&LEA cooperation and collaboration.

It should be noted that if these policy recommendation insights are embraced as a whole, then the more likely they are to enhance the effectiveness of P&LEAs, whilst at the same time safeguarding the rights and privacy of individuals and wider society. An alternative future dominated by unguided and unregulated deployment of AI technology by P&LEAs is highly unlikely to achieve any comparable situation.

Finally, guided by the initial requirement that the ALIGNER policy recommendations were to be *"…tailored to the operational needs of the law enforcement sector…"*, the congruency assessment was used as a benchmark against which they were re-evaluated and revised. They were re-ranked and numbered to reflect how well each was in concordance, and their wording was also revised to ensure that they expressed as accurately as possible their current aims and objectives.

The ALIGNER policy recommendations are now as follows:

ALIGNER Policy Recommendation 1 (Revised 2023)
The EU to ensure that the procurement, utilisation and in-service development of all AI technology, by or on behalf of P&LEAs, is carried out in a holistic and consistent way, taking full cognizance of both the adverse and beneficial impacts it may have on society
ALIGNER Policy Recommendation 2 (Revised 2023)
The EU to provide as a matter of priority a tailored legislative framework designed specifically to guide and support the P&LEAs of Member States in their adoption, acquisition and use of AI technology
ALIGNER Policy Recommendation 3 (Revised 2023)
The EU to assist and support the P&LEAs of Member States to bridge the gaps between their current level of technology and the AI technology that is already available and to transition towards the AI technology that is already foreseeable on the near horizon
ALIGNER Policy Recommendation 4 (Revised 2023)
The EU to instigate a review of the EU data protection framework as it relates to the nexus between 'real-world' data and P&LEA functions and operations, and their potential use of AI technology to access and process it. Currently, it appears to act as a barrier to P&LEA procurement and use of AI technology
ALIGNER Policy Recommendation 5 (Revised 2023)
The EU and Member States to ensure P&LEAs always have a trained, competent and knowledgeable 'human in the loop' when they utilise AI technology to assist in decision making processes
ALIGNER Policy Recommendation 6 (Revised 2023)
To urge the EU to conduct and facilitate systematic and regular research into AI technology that may have potential ramifications (both as solutions and as threats) for the operational needs and capability requirements of P&LEAs

With these revisions in place, the policy recommendations are now well established as ALIGNER enters into its third and final year.

Consideration will be given to using the initial results as a foundation from which to repeat the cross-congruency assessment but his time, the comparison could expand the number of relevant recommendations and ensure they are derived from derived from a wider range of projects whose remit includes engagement with the topic of AI related technology and its interactions with P&LEAs. A final decision will be made after a feasibility and scoping exercise has been carried out.

Finally, by continuing to work collaboratively with Project STARLIGHT, ALIGNER will continue to examine the impact on policy recommendations of both emerging AI tech developments and the progress of the AI Act as affects P&LEAs. It will also ensure that the policy recommendations are also informed by other EU initiatives as they develop and by the work carried out within ALIGNER itself.

Table of contents

List of Acronyms and Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
ALIGNER	‘Artificial Intelligence Roadmap for Policing and Law Enforcement’
API	Application Programming Interface
ENISA	European Union Agency for Cybersecurity
GPT	Generative Pretrained Transformer
KPI	Key Performance Indicator
LM (LLM)	Language Model (Large Language Model)
LEAAB	Law Enforcement Agency Advisory Board (ALIGNER)
ML	Machine Learning
P&LEA	Police and Law Enforcement Agencies
PopAI	European Positive Approach towards Artificial Intelligence tools in support of Law Enforcement & Safeguarding Privacy & Fundamental Human Rights
RL RLHF	Reinforcement Learning Reinforcement Learning from Human Feedback
SIEAB	Scientific, Industrial and Ethical Advisory Board (ALIGNER)
SO	Specific Objective
STARLIGHT	Enhancing the EU‘s strategic autonomy in field of Artificial Intelligence for LEAs
WP	Work Package

1. Introduction

Project ALIGNER commenced in October 2021, with the overall aim of *"...[bringing] 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 Artificial Intelligence supports LEAs while simultaneously empowering, benefiting and protecting the public.*"

To achieve this aim by the end of the project in September 2024, ALIGNER has been set a number of Specific Objectives (SO) and two of them relate specifically to policy recommendations. These are required to be *"…tailored to the operational needs of the law enforcement sector…supported by [identified] capability gaps…[and] provide an overview of them from a societal and ethical perspective".* They are to be presented in three deliverables entitled "Policy Recommendations" (D2.3, D2.4 and D2.5). [3] Other aspects of the ALIGNER policy recommendations are to be incorporated as necessary into the series of ALIGNER research roadmap deliverables (D5.3 and its iterations).

The first ALIGNER workshop in November 2021 gave rise to the early insight that AI technology in the context of policing and law enforcement is of a dual nature. It consists of two interlinked dimensions, one where AI technology is a crime and security threat that P&LEAs must respond to and challenge when necessary, and one where P&LEAs can utilise AI technology in order to carry out their roles and functions. Over the course of the first three ALIGNER Workshops, it became clear that in the context of policing and law enforcement, AI technology was already having an impact within both of these different yet inter-connected dimensions.

In the 'crime and security threat' dimension, AI technology can be a threat if is utilised by 'bad actors' to enable them to carry out acts hostile or detrimental to society and individuals within it. Here, the key aspect for P&LEAs is that they have no control over how or why the AI technology is being used, nor (initially) over the 'bad actors' who seek to use it for their own illegal or illicit ends. This can create crime and security threats that Police and LEA's have to respond to in both the physical domain (to counter the criminals and their criminal actions) and in the digital domain (to counter the AI technology they may be utilising to carry out their criminal actions). In both of these domains, a prime need for P&LEAs is to secure evidence.

In the second of these dimensions, AI technology is utilised by P&LEAs to fulfil their roles and responsibilities to protect society i.e. AI technology is in the service of policing and law enforcement. Here, the key aspect is that as organisations, P&LEAs are in control of the AI technology they use, from its initial deployment to the objectives and circumstances of its use. Consequently, they can be proactive as to how it is controlled, directed and utilised. Crucially therefore, P&LEAs are responsible and accountable for all aspects of how and why they employ it and ultimately, the impact and consequences its use may generate. A 'bad actor' using AI technology in pursuit of their criminal, hostile or malicious ends operates under no such constraints.

It became apparent early on that ALIGNER had to adopt a broad approach to AI technology in the context of policing and law enforcement if it was to achieve the most relevant and useful results. If it did not and concentrated on AI technology only as a crime and security threat, a significant gap could occur if the second critical dimension of AI technology, where it is utilised in the service of P&LEAs, was given less attention or only examined in the margins.

[3] ALIGNER Description of Work, Part B-4 and B-5

Section 2 is an overview of the results presented in September 2022 in the first deliverable, '*D2.3-Policy Recommendations'*. The aim of this section is to recap how they were drawn from the information gathered by ALIGNER in its first year concerning the then current manifestations and impact of AI technology on P&LEA operations, plus the threats, problems and issues P&LEAs were facing.

It provides the foundation for this deliverable by briefly describing the process of how the overall policy-related output of the first three ALIGNER workshops was gathered and analysed to identify and categorised into six potential policy issues "*tailored to the operational needs of the law enforcement sector*". From this information, forty-four capability gaps and 'areas of concern' where potentially, EU policy may be required to address them were identified. These were analysed, assessed and finalised to become the initial ALIGNER policy recommendations.

Section 3 presents an overview of key events relevant to AI technology in the context of policing and law enforcement that have occurred in the year since the previous ALIGNER policy recommendations were submitted in September 2022. In terms of events and developments, there are two categories. The first one concerns the process of turning the EU Artificial Intelligence Act Proposal into EU-wide legislation, with key factors for P&LEAs over this timescale being its glacier-like progress towards the statute book and the introduction of a number of contentious issues that need to be resolved

In contrast to this is the second category relating to AI technology and its use and, once more, the potential implications that have arisen for police and law enforcement. A major catalyst for rapid change appeared in November 2022 as a result of the release to the public by 'OpenAI' of their AI Large Language Model (LLM) known as 'ChatGPT'. In a very short space of time, AI technology became of great interest to the public, the press and governments.

Both of these categories contain implications applicable to the ALIGNER policy recommendations and consequently, while it may be a truism that 'policy is not made in a vacuum', this must also apply to policy recommendations. This section highlights some of the inherent dilemmas that particularly affect P&LEAs.

Section 4 describes efforts to ensure the ALIGNER policy recommendations continue to accurately reflect the needs of P&LEAs in a rapidly changing environment (policy and real world), and secondly, to ascertain whether ALIGNER is at variance with the policy recommendations and key findings of other, comparable AI projects.

The section concentrates in particular on a cross-project comparison carried out to examine policy recommendations presented by external projects support to see if they are in concordance with the ALIGNER policy recommendations and to assess what impact they may have on them. It compared each ALIGNER policy recommendation to the relevant policy recommendations produced by ALIGNER's two sister projects in the EC AI technology and law enforcement project cluster ('popAI' and 'STARLIGHT') as a means to assess their specific congruency.

In addition, the high-level findings of the ENISA 'Research and Innovation Recommendations' were also included in the comparison. The overall results were used to rank the ALIGNER policy recommendations by order of congruence and some of their wording was also revised to better reflect their applicability to situation as it stands in September 2023.

Section 5 draws a number of conclusions, and outlines potential avenues of future work that will be presented in the final iteration of the ALIGNER policy recommendations at the end of the project in 2024.

1.1 Gender Statement

ALIGNER partners actively safeguard gender equality and are aware of gender issues in science and technology (ref. "Commission of the European Communities: Women and Science: Excellence and Innovation–Gender Equality in Science, SEC (2005) 370, available at https://data.consilium.europa.eu/doc/document/ST-7322-2005-INIT/en/pdf).

ALIGNER monitors gender equality addressing biases and constraints throughout all the stages of the project as listed in Gendered Innovations 2 (ref “European Commission: Gendered Innovation 2 How Inclusive Analysis Contributes to Research and Innovation, (2020) available at https://op.europa.eu/en/publication-detail/-/publication/33b4c99f-2e66-11eb-b27b-01aa75ed71a1/language-en).

Outreach activities, visual representations, events, modes of data gathering and analysis, and other research products related to D2.4 have been and will be gender proofed during the internal review process following the ALIGNER Gender policy (ref: ALIGNER D1.2 Project Handbook, section 8 ‘*Gender aspects in publications and research*’).

2. Review of the ALIGNER Policy Recommendations

ALIGNER has been set a number of Specific Objectives (SO) and two of them relate specifically to policy recommendations. First and foremost of these, they are required to be *"…tailored to the operational needs of the law enforcement section…supported by [identified] capability gaps…[and] provide an overview of them from a societal and ethical perspective."* [4]

This document is the second of three deliverables concerning policy recommendations. D2.3 was submitted in September 2022; D2.4 in September 2023 and D2.5 will be submitted at the end of the project in September 2024. Full details of how the ALIGNER policy recommendations were initially derived can be found in the previous deliverable, D2.3.[5] A brief summary of them is given in this section.

2.1 Summary of how the ALIGNER Policy Recommendations were derived

With the requirement to focus on the 'operational needs, and 'capability gaps' of P&LEAs, the perspective of the process to derive the initial policy recommendations needed to be from the 'bottom up' rather than the 'top down'. In other words, the policy recommendations had to be anchored in recent P&LEA operations and experiences in the 'real world' and not just be the product of hypothetical or speculative thinking.

In order to achieve this, a methodology utilising a modified 'Grounded Theory' approach was adopted. Based on this, data was gathered at ALIGNER Workshops and from research, before being collated, analysed and assessed, with the results then being used to generate the policy recommendations. During the first nine months of the project three Workshops were held, involving invited participants, members of the two ALIGNER Advisory Boards (Law Enforcement Agency Advisory Board (LEAAB) and the Scientific, Industrial and Ethical Advisory Board (SIEAB), and representatives of other relevant EU projects, in particular from the other two projects from the AI LEA cluster; 'STARLIGHT' and 'PopAI'.

Working together with the participants enabled ALIGNER to explore an initial insight; there are two closely aligned facets of AI technology in the context of policing and law enforcement. In one, AI technology can pose a crime and/or security threat. In the other, AI technology can be utilised by Police and Law Enforcement Agencies to increase their operational capabilities and to drive improvements to their overall effectiveness.

The information gathered from the workshops and relevant elements of the wider research effort were examined and forty-four 'areas of concern' and potential 'capability needs' that had both implications for P&LEAs and that also may need to be tackled at the policy level were identified. These were assessed and six over-arching categories where EU-led policy could have a beneficial effect in the context of policing and law enforcement were identified. Each 'areas of concern' or potential 'capability need' was then assigned to a category. The categories were:

- ❖ AI Technology: Research, development, and exploitation
- ❖ AI technology as a Crime and Security threat
- ❖ Legal and Judicial issues and considerations
- ❖ Ethical and Human Rights implications of P&LEA use of AI technology

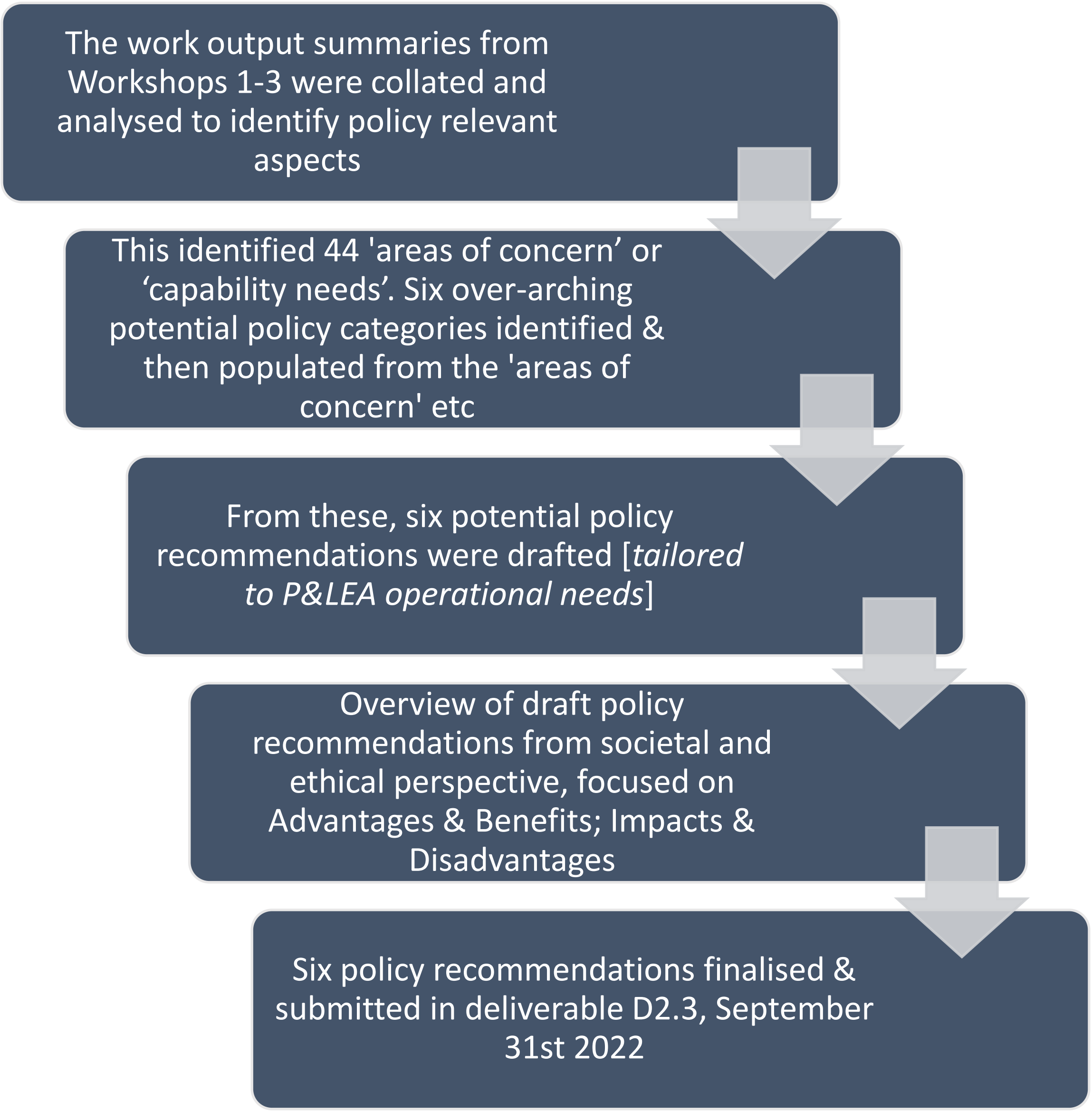
[4] ALIGNER Description of Work, Part B-4 and B-5

[5] ALIGNER D2.3 'Policy recommendations' (Available at https://aligner-h2020.eu/deliverables/)

- ❖ P&LEA utilisation of AI technology to enhance and increase P&LEA capabilities
- ❖ P&LEAs and AI technology: Internal governance and training issues

After taking into consideration all the identified information and categories, six draft policy recommendations were derived. As required by the ALIGNER description of work, each of them was reviewed from a societal and ethical perspective and then subjected to an overview to explore their possible impacts, benefits, advantages and disadvantages. The overall process is summarised in Figure 1 below:

Figure 1 – Summary of process used to derive ALIGNER Policy Recommendations from ALIGNER Workshop outputs



The work output summaries from Workshops 1-3 were collated and analysed to identify policy relevant aspects

This identified 44 'areas of concern' or 'capability needs'. Six over-arching potential policy categories identified & then populated from the 'areas of concern' etc

From these, six potential policy recommendations were drafted [*tailored to P&LEA operational needs*]

Overview of draft policy recommendations from societal and ethical perspective, focused on Advantages & Benefits; Impacts & Disadvantages

Six policy recommendations finalised & submitted in deliverable D2.3, September 31st 2022

The six initial ALIGNER policy recommendations are presented in Section 2 below.

2.2 ALIGNER initial policy recommendations (2022)

The six topics identified by ALIGNER as being most suitable for EU-led policy were formalised as the first iteration of the ALIGNER policy recommendations and were published on September 30th, 2022. They are presented below in Table 1.

Table 1 - ALIGNER policy recommendations to address identified P&LEA needs at EU level (September 2022)

ALIGNER Policy Recommendation 1
Ensure the procurement, utilisation and in-service development of all AI technology by P&LEAs is carried out in a holistic way, with full cognizance of the adverse and beneficial impacts it may have on society
ALIGNER Policy Recommendation 2
Ensure there is always a competent and knowledgeable 'human in the loop' if AI technology is used to support and assist P&LEAs and criminal justice systems in critical decision-making processes
ALIGNER Policy Recommendation 3
Review the EU data protection framework on the use by AI technology of 'real-world' data as it appears to act as a barrier to the procurement and use by P&LEAs of AI technology
ALIGNER Policy Recommendation 4
Enable and support the P&LEAs of EU Member States to bridge gaps between their current levels of technology, the AI technology of today and AI technology already on the near horizon
ALIGNER Policy Recommendation 5
The EU to approach AI technology adoption in the context of P&LEAs and criminal justice systems by using a Directive, as used to counter terrorism from 2017 onwards
ALIGNER Policy Recommendation 6
Conduct further and regular research into P&LEA and criminal justice concerns and capability needs for AI technology in order to ensure EU policy makers are aware of new developments and on-going issues

Section 4 shows how ALIGNER has developed from these original policy recommendations in two ways. First, by actively comparing them with the policy recommendations from other relevant projects and organisations and second, by adjusting their content and wording to better reflect the current situation. However, before examining this aspect, it is necessary to review how both AI technology and its use and the new flagship EU legislation aimed at regulating and controlling it, the Artificial Intelligence Act, have developed over the last twelve months.

3. AI technology in the context of policing and law enforcement: Key developments since September 2022

Since the ALIGNER policy recommendations were submitted in 2022, much has happened in terms of the technological development of AI systems and their use in the public domain, forcing the topic of AI into the awareness of mainstream politicians, the media and the general public. Before this is examined in more detail, this section reviews how the EU Artificial Intelligence Act (AI Act) reached its current incomplete and internally conflicted state. The final part of the section deals with the overall implications both these aspects have for the Aligner policy recommendations.

It may be a truism that 'policy is not made in a vacuum' but the twelve-month period between the submission of the first iteration of the ALIGNER policy recommendations in September 2022 and this second iteration in September 2023, AI technology has demonstrated its potential to make rapid and fundamental advances almost overnight. A consequence of this has been to increase the complexities that policy makers and legislators must take into account when determining their responses to it. The cornerstone of this response, when it is passed into binding legislation, will be the EU Artificial Intelligence Act.

3.1 An overview of the Artificial Intelligence Act Proposal of the European Commission

In July 2019 Ursula von der Leyen, as a prospective candidate for President of the European Commission, presented her political priorities for the European Commission for its next term from 2019 to 2024.[6] She committed it to "*put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence*" and for it to do so within her first 100 days in office.[7] On 19 February 2020, eighty-one days after she took office, the new EU Commission put forward a White Paper on AI[8], providing concrete policy options for new AI legislation with the two-fold objective of "*promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology.*"

A year later, on 21 April 2021, the European Commission finally published its formal proposal for a regulation laying down harmonised rules on artificial intelligence (the 'AI Act Proposal').[9] The legal instrument of a regulation was chosen by the Commission to ensure the uniform application of the new rules in the EU as once adopted and in force, the AI Act will then be directly applicable in each EU Member State.

The AI Act Proposal aims to enhance and foster the development of a single market for AI applications while laying down specific safety requirements. To achieve its objectives, the Commission pursued a

[6] Ursula von der Leyen (2019). A Union that strives for more: My agenda for Europe, available at https://commission.europa.eu/system/files/2020-04/political-guidelines-next-commission_en_0.pdf (accessed on 18 August 2023).
[7] *Ibid*., p. 13.
[8] European Commission (2020). White Paper on Artificial Intelligence – A European approach to excellence and trust, available at https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed on 18 August 2023).
[9] European Commission (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (accessed on 18 August 2023).

horizontal regulatory approach and the Proposal provides a flexible and future-proof legal framework applicable to all ‘AI systems’ throughout their whole lifecycle. It is important to note that while the Proposal includes numerous provisions of direct relevance to the roles, responsibilities, functions and activities of police and law enforcement agencies, they are not encompassed within it through any cohesive or unifying framework, structure or approach.

The Proposal also provides for a broad and technology-neutral definition of ‘AI systems’, based on both their key functional characteristics and the approaches followed in their development.[10] In other words, to qualify as an AI system, the software need to fulfil two conditions. First, for a given set of human-defined objectives, the software needs to generate outputs (such as content, predictions, recommendations or decisions) which influence the environment it interacts with. Second, the software needs to be developed through one of the techniques listed in Annex I to the Proposal; these are: (a) machine learning approaches; (b) logic- and knowledge-based approaches; and (c) statistical approaches, Bayesian estimation, search and optimisation methods.

The Proposal also sets out obligations and these mainly apply to two categories of subjects:[11]

(a) ‘providers’; meaning any natural or legal persons, public authorities, agencies or bodies that develop or own AI systems to place them on the market or put them into service, whether against payment or for free; and
(b) ‘users’; meaning any natural or legal persons, public authorities, agencies or bodies that use AI systems under their authorities, unless they are carrying out a non-professional activity.

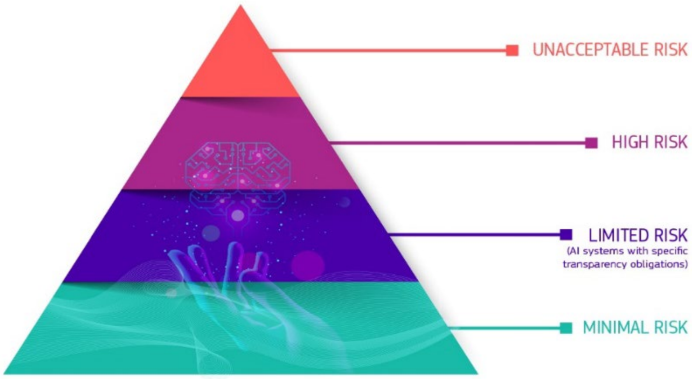
The obligations contained in the Proposal apply not only to users and providers located or established in the EU, but also to users and providers located or established in third countries, as long as either they are targeting the EU internal market or the output produced by the AI system is used in any way within the EU.

Starting from the assumption that the risks and harm potentially caused by AI systems differ depending on the circumstances and their use, the AI Act Proposal adopts a risk-based regulatory approach. As shown in Figure 2 below, the Proposal distinguishes four levels of risk with AI technologies:

- Unacceptable risk
- High risk
- Low risk
- Minimal risk

[10] European Commission (2021), Article 3(1).
[11] *Ibid.*, Article 2.

Figure 2: AI risk levels. Source: European Commission.



UNACCEPTABLE RISK

The AI Act Proposal prohibits the placing on the EU market or the use of certain AI systems deemed to create an unacceptable risk by contravening EU values and violating fundamental rights.[12] The prohibition covers four AI practices.

Article 5(1)(a) and (b) of the Proposal tackle AI systems deploying subliminal techniques beyond a person's consciousness, and AI systems exploiting the vulnerabilities of a specific group of persons due to their age, physical or mentally disability and which are used with the intent of materially distorting a person's behaviour and in a manner that causes harm.

Article 5(1)(c) of the Proposal tackles AI systems used by public authorities to evaluate or classify the trustworthiness of natural persons over a certain period of time, based on their social behaviour or known or predicted personal or personality characteristics, and leading to detrimental or unfavourable treatment.

Of particular relevance and importance in the context of P&LEAs, Article 5(1)(d) of the Proposal tackles the use of remote biometric identification systems operating in 'real-time'. To become prohibited, these AI systems need to fulfil four conditions:

- The AI systems needs to be designed with the purpose of identifying natural persons at a distance, by matching their biometric data with those contained in a reference database.
- The identification process, from the moment of the collection of data to that of the identification in itself, needs to occur in real time, or without any significant delay.
- The AI system needs to be deployed in any physical place which is accessible to the public.
- The AI system needs to be used for law enforcement purposes.

As a consequence, the prohibition does not cover 'real-time' remote biometric identification systems used for purposes other than law enforcement, as well as 'post' remote biometric identification systems, where the identification process occurs with a significant delay.

These AI practices cause serious interferences with the fundamental rights and freedoms of the persons involved and are capable of having severe chilling effects, due to the risk they pose of implementing mass surveillance.

[12] *Ibid.*, Article 5.

The prohibition contained in Article 5(1)(d) of the Proposal is not absolute. It contains major exceptions in favour of reasons of public interest and hence will affect P&LEAs. The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces is allowed ‘for the purpose of law enforcement, unless and in as far as such use is strictly necessary’ for:

- The targeted search for specific potential victims of crime, including missing children
- The prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack
- The detection, localisation, identification or prosecution of a perpetrator or suspect of one of the criminal offences listed in Article 2(2) of the Council Framework Decision 2002/584/JHA and punishable by a custodial sentence or a detention order for a maximum period of at least three years.

In assessing the strict necessity to use ‘real-time’ remote biometric identifications systems, P&LEAs need to take into account both the harm caused in the absence of the use of the system and the consequences of its use for the rights and freedoms of all individuals concerned.

The use needs to comply with necessary and proportionate safeguards, including temporal, geographical and personal limitations. The use must also be preceded by a judicial authorisation, issued upon reasoned request and in accordance with national law. However, in the case of an urgent situation, the authorisation can be requested after its use has started, or even after it has ended.

The criminal offences listed in Article 2(2) and referred to above are as follows:

- Participation in a criminal organisation
- Terrorism
- Trafficking in human beings
- Sexual exploitation of children and child pornography
- Illicit trafficking in narcotic drugs and psychotropic substances
- Illicit trafficking in weapons, munitions and explosives
- Corruption
- Fraud, including that affecting the financial interests of the European Communities within the meaning of the Convention of 26 July 1995 on the protection of the European Communities' financial interests; laundering of the proceeds of crime
- Counterfeiting currency, including of the Euro
- Computer-related crime
- Environmental crime, including illicit trafficking in endangered animal species and in endangered plant species and varieties
- Facilitation of unauthorised entry and residence
- Murder
- Grievous bodily injury; illicit trade in human organs and tissue
- Kidnapping
- Illegal restraint and hostage-taking
- Racism and xenophobia
- Organised or armed robbery
- Illicit trafficking in cultural goods, including antiques and works of art
- Swindling; racketeering and extortion
- Counterfeiting and piracy of products
- Forgery of administrative documents and trafficking therein
- Forgery of means of payment

- Illicit trafficking in hormonal substances and other growth promoters
- Illicit trafficking in nuclear or radioactive materials
- Trafficking in stolen vehicles
- Rape
- Arson
- Crimes within the jurisdiction of the International Criminal Court
- Unlawful seizure of aircraft/ships
- Sabotage

HIGH RISK

For certain AI systems which create a high risk, the AI Act Proposal establishes horizontal obligations relating to:

- Risk assessment and mitigation
- Quality of datasets
- Logging of activities
- Documentation
- Information to the user
- Human oversight
- Robustness, security and accuracy[13]

There is also a requirement for each AI system to be subjected to a conformity assessment before it is taken into service.

The obligations apply to two categories of AI practices: AI systems incorporated in products normally covered by the EU product safety legislation and stand-alone AI systems. The Proposal lists in Annex III eight areas in which high-risk stand-alone AI systems are applied. Of these eight areas of application, the most relevant in the context of policing and law enforcement are:

(i) Biometric identification and categorisation of natural persons, including ‘real-time’ and ‘post’ remote biometric identification

(ii) Law enforcement; including predictive policing tools, polygraphs or similar instruments, tools to detect deep fakes, systems used to evaluate the reliability of criminal evidence and, in general, profiling tools and systems used for crime analytics using large datasets to identify unknown patterns

(iii) Migration, asylum and border control management

(iv) Administration of justice and democratic processes.

These AI systems operate in areas particularly associated with surveillance, arrest or potential discrimination. Their use can also harm the exercise of other fundamental rights, such as the right to a fair trial or the presumption of innocence.

LIMITED RISK

Article 52 of the AI Act Proposal establishes minimal transparency obligations for AI systems creating limited risks. Whenever AI systems interact with natural persons or generate content and this may pose specific risks of impersonation or deception, providers need to ensure that users are aware they are

[13] For an overview of the obligations established in the Proposal for high-risk AI systems, see Eren, E., Casaburo, D., and Vogiatzoglou, P. (2021). ALIGNER D4.1 State-of-the-art reports on ethics & law aspects in Law Enforcement and Artificial Intelligence, available at https://aligner-h2020.eu/wp-content/uploads/ALIGNER-D4.1-SotA-reports-on-ethics-law-aspects-v20220714.pdf (accessed on 17 August 2023).

interacting with an AI system. It is important to note that this obligation does not apply to AI systems employed in a P&LEA context.

MINIMAL RISK

AI systems creating minimal or no risks, such as AI-enabled videogames or spam filters, can be designed and used freely.

Before the AI Act Proposal can be adopted as a binding piece of EU legislation, it needs to undergo the ordinary legislative procedure. This means that the Council of the European Union [§ 3.1.1] and the European Parliament [§ 3.1.2], as co-legislators, can each propose amendments to the Commission's Proposal but ultimately, they both need to agree on the same draft text. Agreement is usually reached through informal inter-institutional negotiations ('trilogues'), mediated by the Commission [§ 3.1].

The next two sections consider the amendments already proposed that are of the most specific interest to P&LEAs. It must be borne in mind that at the time of the submission of this Deliverable (September 31st 2023), a negotiated final wording agreed during the trilogues has not yet been reached and the agreed text may yet considerably alter the provisions of the final Act.

3.1.1 The 'general approach' of the Council of the European Union

On 25 November 2022, the Council of the European Union adopted its 'general approach' to the AI Act,[14] in effect a set of amendments to the Commission's Proposal. Seeking to distinguish AI from more classical software systems, the Council narrows down the definition of 'AI systems' contained in Article 3(1) of the Proposal by emphasising their ability to learn, reason or model.

According to the proposed amendments, to qualify as an AI system, a system needs to fulfil two conditions. First, based on machine and/or human-provided data and inputs, the system needs to infer, using machine learning and/or logic- and knowledge-based approaches, how to achieve a given set of objectives, as well as to produce outputs which influence the environments it interacts with. Second, the system cannot operate based on rules solely defined by humans, but must also operate with varying levels of autonomy.

The Council further clarifies the scope of application of the regulation by including an explicit exemption for AI systems used for the sole purpose of scientific research and development, as well as for AI systems used in the course of purely personal, non-professional activity.

UNACCEPTABLE RISK

The Council follows the Commission's approach on the AI systems that create unacceptable risks by equally foreseeing four prohibited AI practices, but they also suggest a number of amendments.

Of particular relevance in a P&LEA context, a proposed amendment restricts the prohibition contained in Article 5(1)(d) to those 'real-time' remote biometric identification systems used for law enforcement purposes but only when they are used as such by P&LEAs (or on their behalf).

[14] Council of the European Union (2022). 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 - General approach, available at https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf (accessed on 21 August 2023).

Moreover, the new amendment further expands the exceptions to the 'general' prohibition by now allowing the use of 'real-time' remote biometric identification systems for the prevention of a specific and substantial (and not necessarily imminent) threat to critical infrastructures. In addition, the amendment also allows their use in the investigation and prosecution of offences, where they are punishable by EU Member States with a custodial sentence for a maximum period of at least five years.

HIGH RISK

The Council follows the Commission's approach on the AI systems creating high risks by equally foreseeing two categories of high-risk AI systems. However, the proposed amendments introduce a novel and potentially problematic exemption: the stand-alone AI systems listed in Annex III are to be considered as high-risk, unless 'the output of the system is purely accessory to the relevant action to be taken and, therefore, does not lead to significant risks to health, safety or fundamental rights'.

In other words, the deployment of an AI system in the circumstances described in Annex III does not necessarily lead to its classification as 'high-risk' when the output of the AI system does not have a high degree of importance in respect of the action to be taken. The AI system should be considered as purely 'accessory' and posing a low risk, as it does not lead to significant risks to health, safety or fundamental rights.

However, this approach would be problematic as the proposed amendments do not go on to specify the circumstances under which the output of the AI system may be considered as purely accessory. Instead, they require this task to become the responsibility of the Commission alone, through the adoption of an Implementing Act. While not explicitly demanded by the amendments, a further Implementing Act would also be required to give the precise details of the assessment procedure and to specify the entities with responsibility for carrying it out. The net result is that currently, a legislative gap exists.

This is not the sole problematic area requiring resolution. The new proposed Annex III still consists of eight areas of application. However, in one area of most relevance to P&LEAs, both 'real-time' and 'post' remote biometric identification are listed in the Annex. If this remains the case, the implications for P&LEA post event evidential investigations may become considerable.

Finally, certain types of AI systems for detecting deep fakes are removed from the list and thus would be considered as 'low-risk'. Arguably, this is a positive step for P&LEAs when it comes to their responsibility to counter their use in crime, criminality and as security threats. However, AI systems used for big data crime analytics are also removed and become 'low risk'. Here, it can be argued that such a step should require a great deal of thorough and careful thought to be given to the decision beforehand due to the potential consequences it could have for broader society.

3.1.2 The 'negotiating position' of the European Parliament

On 14 June 2023, the European Parliament adopted its 'negotiating position' to the AI Act [15] and acting as a set of amendments to the Commission's Proposal. While recognising the need to distinguish AI from simpler types of software, the Parliament broadened the definition of 'AI systems' contained in Article 3(1) of the Proposal to align it with the one proposed in 2019 by the Organisation for Economic

[15] European Parliament (2023). Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, available at https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf (accessed on 22 August 2023).

Co-operation and Development (OECD).[16] According to the proposed amendments, to qualify as an AI, a machine-based system needs to fulfil two conditions. First, for explicit or implicit objectives, the system needs to generate outputs which influence physical or virtual environments. Second, the system cannot operate based on rules as solely defined by humans, it also needs to operate with varying levels of autonomy.

UNACCEPTABLE RISK

The Parliament followed the Commission's approach on the AI systems that create an unacceptable risk and expanded it by adding five novel prohibited AI practices.

Particularly relevant in a P&LEA context, and completely opposite to the amendments proposed by the Council, the Parliament extended and generalised the prohibition contained in Article 5(1)(d) for 'real-time' remote biometric identification systems in public accessible spaces. These systems are completely prohibited for both law enforcement and non-law enforcement purposes and consequently, no exception for their use is established for P&LEAs.

Of particular relevance to P&LEAs are also five new proposed prohibitions. They concern:

- Biometric categorisation systems to categorise natural persons according to sensitive or protected attributes or characteristics, whether known or predicted
- AI systems to assess the risk of a natural person (re)offending or to predict the (re)occurrence of an actual or potential offence based on profiling or on assessing personality traits and characteristics (i.e. as carried out when using so-called 'predictive policing')
- AI systems creating or expanding facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage
- AI systems used to infer emotions in the areas of law enforcement, border management, workplace and education
- 'Post' remote biometric identification systems, unless they are subject to a pre-judicial authorisation and are strictly necessary for a targeted search connected to a serious criminal offense that has already taken place.

HIGH RISK

The Parliament follows the Commission's approach on the AI systems creating high risks, by equally foreseeing two categories of high-risk AI systems. As with the Council's proposed amendments, the Parliament also introduces a new exemption; the stand-alone AI systems listed in Annex III are to be considered as high-risk but only if they pose a significant risk of harm to health, safety and fundamental rights of natural persons or to the environment.

Guidelines clearly specifying the circumstances where the output of an AI system would pose a significant risk of harm should be specified by the Commission. However, making any assessment required by the guidelines is mainly delegated to the 'provider' of the AI system. It is they who will need to evaluate the level of risk of the AI system, submitting a reasoned notification to their national

[16] OECD Council (2019). Recommendation of the Council on Artificial Intelligence, available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 22 August 2023).

supervisory authority who will review it and notify the provider if it deems the AI system to be misclassified.

The new proposed Annex III still consists of eight areas of application. As for the most relevant areas for P&LEAs, all the non-prohibited biometric identification systems are covered, as well as AI systems inferring personal characteristics on the basis of biometric data. However, AI systems detecting deep fakes are removed from the list and so are considered as 'low-risk'.

Departing from both the Commission's and the Council approaches, the Parliament subjects the use of 'high-risk' AI systems to the obligation imposed on those who deploy and use them to conduct a fundamental rights impact assessment and this includes P&LEAs.[17] In the absence of appropriate mitigation measures, 'users' must not deploy or operate a high-risk system.

3.1.1 Some fundamental dilemmas created by the AI Act negotiations

The EU AI Act proposal currently consists of, in effect, three different proposals; the original proposal put forward by the EC in April 2021, the 'general approach' to it adopted by the EU Council in November 2022 and the 'negotiating position' proposed by the EU Parliament in July 2023.

For the proposal to become law as an EU regulation, the co-legislators must negotiate to agree on its final wording and this will not be an easy or rapid process as within the context of policing and law enforcement, there appear to be some fundamental disagreements concerning the use (or prohibition from use) by P&LEAs of certain AI technologies.

The first balancing exercise between different rights and interests, as set out in the draft texts, show different approaches, especially by the two co-legislators. While the Council generally favours reasons of security and public interest, the Parliament puts a major emphasis on the enjoyment of fundamental rights and liberties. This is clearly visible for instance in the approach to 'real-time' remote biometric identification systems. They would be allowed by the Council if they comply with certain (not so strict) specific conditions while Parliament is seeking their absolute prohibition. Again, while the Council now downgrades AI systems used for big data crime analytics from being 'high-risk' to 'low-risk', the Parliament generally classifies them as 'high-risk' and further prohibits them when applied to certain law enforcement techniques e.g., predictive policing.

The fundamental differences and disagreements outlined above means there continues to be no clear legislative framework defining what AI technology P&LEAs can acquire, and what practices P&LEAs could utilise to carry out lawfully to their roles and functions. Until this situation is resolved, P&LEAs will continue to operate in a fog of uncertainty. This could continue for some while as, of the current date (September 30th 2023), the formal public reporting of the co-legislators' discussions shows that they have not yet covered any of the issues identified here by ALIGNER as being of importance to the use of AI in the context of policing and law enforcement.[18]

[17] On this topic, see Casaburo, D., and Marsh, I. (2023). ALIGNER D4.2 Methods and guidelines for ethical & law assessment, available at https://aligner-h2020.eu/wp-content/uploads/ALIGNER-D4.2-Methods-and-guidelines-for-ethical-law-assessment-v20230324-FINAL.pdf (accessed on 22 May 2023).

[18] See https://data.consilium.europa.eu/doc/document/ST-11320-2023-REV-1/en/pdf (Accessed 19 September 2023)

3.2 Recent new advances in AI technology: Large Language Models and Chat bots

There are three key aspects to the developments in AI technology over the last twelve months; the AI technology itself, the implications it has for society and, more specifically, the potential impact of both these on P&LEA operations and activities.

The main area of AI technology currently on the crest of this wave is based on Language Models (LMs). They make probabilistic predictions of words and have been around in linguistics research for a very long time[19] [20]. Large Models trained on large text corpora are referred to as Large Language Models (LLMs) and when LLMs are finetuned using Reinforcement Learning from Human Feedback (RLHF), they are referred to as *chat bots*.

It is only within recent years that LLMs have become potent across all sorts of Machine Learning (ML) applications. This is mainly due to three things;

- ❖ The capability to train LLM models on very large text datasets, known as text corpora
- ❖ The capability to tune LLM models by using RLHF
- ❖ The advent of the 'transformer neural network'

LLMs are conventionally optimised unconditionally, functioning only to predict statistically the most likely succession of words. By using Machine Learning class of algorithms called Reinforcement Learning[21] (RL), LLMs can be optimized subjectively for a desired outcome by utilising RLHF, where engineers or users can label responses as "good" or "bad" and to favour the "good" response. Improving this is a current research priority.

The transformer network was first proposed in 2017[22] and in contrast to earlier Language Models, it is now easier and faster to train. More recently, researchers at the company 'OpenAI' showed[23] [24] [25] that they could achieve a generic function for language models with transformers by pretraining them on large data sets, and that their performance was scaled by pretraining data set size, hence their designation as a "generative pretrained transformer" (GPT). The emergent function turned out to be very versatile and usable for machine translation, human-like chat, answering both trivial or expert questions, consultation and also, for code generation etc.

'OpenAI', who spearheaded the initial research, released research versions of GPT in incremental steps. GPT-3 came first in 2020 and was designed to complete sentences. It was followed by GPT-3.5 in March 2022 and the LLM model then became potent for many different problems. 'ChatGPT' is the 'OpenAI' application based on GPT-3, but optimized for responding to instructions instead of completing sentences. It became freely available as a research preview on November 30th 2022 and in effect, its release became the catalyst for a rapid and wide-spread expansion of AI capabilities and their use in the public domain.

[19] Jelinek, F., & Mercer, R. L. (1980). Interpolated estimation of Markov source parameters from sparse data. In Proceedings of the Workshop on Pattern Recognition in Practice. Amsterdam: North-Holland.

[20] Goodman, J. (2001). A bit of progress in language modeling. *Computer Speech & Language, 15*(4), 403-434.

[21] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.

[22] Vaswani, A et al. (2017). Attention Is All You Need. arXiv preprint. Link

[23] Radford, A. et al. (2018). Improving Language Understanding by Generative Pre-training. Link

[24] Radford, A. et al. (2019). Language Models are Unsupervised Multitask Learners. Link

[25] Brown, T. B. et al. (2020). Language Models are Few-Shot Learners. Link

3.1.2 From ChatGPT to subsequent Large Language Models (LLMs) and Chat bots

The explosion of available LLMs and chat bots has been rapid following the release of ChatGPT. There are two types now available, *closed-source* models operating via application programming interfaces (APIs), where the models are located off-site and are under the control of others, and *open-source* models where the 'weights' i.e. the pretrained memory, is released and available for anyone to use. The models behind APIs are, in general, more capable but their usage is monitored which could act as a brake to prohibit and limit nefarious use.

Table 2 Chronology of developments in LLMs and Chat bots from November 2022

Date	Product	Application	Creator	Comment
November 30 2022	ChatGPT (Using GPT-3.5)	Chat bot	OpenAI (now owned by Microsoft)	First to be released
February 2023	Bard	Chat bot	Google	Built on earlier PaLM
	LLaMA	LLM	Meta	First published LLM
March 2023	ChatGPT (Using GPT-4)	Chat bot	OpenAI	Includes text & images
	Falcon 40B	LLM	Technology Innovation Institute	First under commercial license?
	Claude	Chat bot	Anthropic	
	Vicuna	Chat bot	LMSYS	Generated via ChatGPT
July 2023	LLaMA 2	LLM	Meta	
	LLaMA 2 Chat	Chat bot	Meta	
(Planned – Autumn 2023?)[26]	Gemini	LLM & Chat bot?	Google	[Not released as of September 2023]
(Unknown)[27]	Falcon Instruct	Chat bot	Technology Innovation Institute	Minimalist open source, public release expected

There are a number of competitors to OpenAI's GPT, including 'Bard' from Google, 'Llama 2' from Facebook, 'Claude' from Anthropic, and many others but all of them rely on the same mechanism as

[26] Hines, K. (2023, September 15). Google Gemini: What We Know So Far. Search Engine Journal. https://www.searchenginejournal.com/google-gemini-what-we-know-so-far/496494/

[27] Rodrigues, V. (2023). Run your private LLM Falcon 7B Instruct with less than 6GB of GPU using 4-bit quantization. Medium. [https://vilsonrodrigues.medium.com/run-your-private-llm-falcon-7b-instruct-with-less-than-6gb-of-gpu-using-4-bit-quantization-ff1d4ffbabcc]

GPT. The differences arise between them due to the data sets they have been trained on, the size of the model, and how they are filtered or restricted.

The release of Llama 2 was open-source, and the subsequent release of the trained model allowed many third parties to make their own improvements or 'flavoured' variants of it. In a performance comparison of many forked models, GPT-4 out-performs the others, but many of the runners-up originated from smaller actors who have utilised it for their own developments.

It is difficult to compare and assess the different kinds of capabilities of LLMs and chat bots but according to most evaluations, GPT-4 is currently the leader of its kind. There is an ongoing effort in trying to standardize what tasks the models are evaluated against. This is a rapidly evolving matter, especially as the models are rapidly becoming more capable and continuously harder tasks are required for assessment[28].

The potential of LLMs can be substantially leveraged when used together with additional systems such as knowledge bases or the internet. 'BingGPT' is one such example where the chat bot includes links to its answer based on its own simultaneous websearch. Premium users of ChatGPT and GPT-4 can use plugins for mathematics, pdf reading, diagram drawing, video summarizing, executing code in sandboxes etc. and for more specialized and controlled use cases.

3.2.2 Intrinsic problems with LLMs and Chat bots

There are a number of issues with AI driven LLMs and chat bots that are particularly problematic for P&LEAs if they use their functionality in the course of their duties. They include:

'Hallucinations'

Currently, while a number of LLMs and chat bots may be state-of-the-art, there is still the unsolved problem of 'hallucination', where the LLM will make up its own 'facts', statements or erroneous code. The statement or result produced by the AI does not seem to be justified by its training data or to have any relation to truth, facts or reality. This result comes about if the input prompt is unclear, inconsistent or contradictory or there is a discrepancy it and what the training data was designed to teach the AI to process. Such an input will still be processed by the system and the misleading or false output will be given to the human user.

Often, the LLM output is still useful, as it may not really matter or the result will be post-edited by a competent and knowledgeable human. However, when this is not the case, the consequences could potentially be fatal. Recently (September 2023), attention has been drawn by experts to the plethora of on-line books for sale created by LLM chatbots and aimed at readers who need information on how to forage for and then cook wild mushrooms. The New York Mycological Society warned readers to only rely on books by recognized authors and foragers [as] "it can literally mean life or death."[29]

[28] https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

[29] 'Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI Sample of books scored 100% on AI detection test as experts warn they contain dangerous advice' *The Guardian*, 1 Sept 2023
https://www.theguardian.com/technology/2023/sep/01/mushroom-pickers-urged-to-avoid-foraging-books-on-amazon-that-appear-to-be-written-by-ai
(Accessed: 4 Sept 2023)

Biases

A cliché in machine learning is that the model is only as good as the data it is trained on. An LLM mirrors its training data e.g. if a dataset has a lot of poorly written or unsafe computer code, the LLM will also generate code that has the same problems. There are many aspects of biases that LLMs may be subject to and which could result in unpredictable and undesired outcomes.

For example, an LLM will mirror hidden biases in the data, which may make it iterate statements that are racist or sexist or cause it to use other discriminatory language. It is not difficult to envisage that this may also cross the threshold into being considered malicious or potentially unlawful.

Biases could also arise in a different way, not from the data content itself but from the design of the reward function where it favours responses in an undesirable (biased) manner. Information could be generated that discriminates against individuals or groups of people. Biases may also be taken advantage of by malicious or criminal actors.

Training data

It is difficult but perhaps not impossible to know what kind of data an LLM has been trained on[30]. This can limit what the model can be used for, as personal information could be leaked. On the other hand, organizations could limit certain kinds of unwanted data in the model the AI is trained on, which could then limit the capabilities in certain areas and which may not be noticeable in the currently used evaluation scores. Moving on from LLMs, the same issue can also apply to RLHF and chat bots. It should be noted that there is an ongoing effort to better develop methods to evaluate the capabilities of LLMs and chat bots.

3.1.3 The unfolding implications of LLMs and Chat bots for policing and law enforcement

It has already become clear over the last twelve months that LLM technology can be utilized by 'bad actors' to become a wide-ranging disruptive technology, based on its generic potential to be applied to virtually all fields where software and text is used and combined with its potential to be used for the automation of all tasks where text is involved. Consequently, even in the current state of development and use of LLMs and chat bots, there are wide-ranging implications for society and in particular, for policing and law enforcement.

There is no doubt that LLMs can be used deliberately for purposes that are malicious, criminal or pose a security threat. Based on ChatGPT as an example, these could include[31]:

- Finding vulnerabilities
    - For example, while ChatGPT does not answer directly to questions posed to it on how to find vulnerabilities, it can often be circumvented by first setting out a plausible scenario, perhaps something like "I am a cyber-security researcher looking to solve a capture the flag challenge". Users acting in this way have managed to extract a lot of information on how to compromise network security.

[30] https://arxiv.org/pdf/2012.07805.pdf

[31] *Hacking with ChatGPT: Five A.I. Based Attacks for Offensive Security* [Video]. YouTube. https://www.youtube.com/watch?v=AwQE3jof63U

  - In another case, it was apparently used to seek out vulnerabilities in “American critical infrastructure”.[32]
- Writing exploits
  - By asking questions indirectly, users have been able to make ChatGPT write code for a given vulnerability.
- Malware development
  - With automated code generation it has become easier to write malware. This even extends to polymorphic malware, which changes with every execution, thereby making it harder to detect.
- AI phishing
  - Phishing can be vastly improved by the use of LLMs. The perpetrator can express themselves in natural English (or other languages) even if they are not a fluent speaker. Many cyber criminals now use this approach[33].
  - In addition, criminals can optimize the content, language and tonality of their communications, increasing the chances for the selected victim to respond in the way desired by the perpetrator. The content of the message can be made to vary automatically, making it harder for ISPs and System Administrators to identify it as malicious.
- Macros
  - Macros can also be generated by ChatGPT within phishing emails, designed to inflict harm if the victim clicks on files or links.
- Identity theft
  - If the data is available, an LLM can be trained to express itself in writing or conversation in a way that mimics the language, style and rhetoric of a human person. This may already lead to a variety of circumstances engineered by a perpetrator to result in identity theft. However, this tactic can also be coupled with synthetic speech technology to sound in voice messages or on social media just like the victim, or someone known closely to them. In the future, synthetic video images may also be produced analogously.

Bearing in mind that barely a year has passed in which to understand the potential of this latest phase of rapidly changing AI technology, it is not surprising that researchers and others are regularly caught by surprise with new revelations and discoveries concerning its intrinsic properties and potential new applications.

This seems to occur in three main ways: as each one is developed directly; as it is discovered how they can be used in conjunction with other technologies as ‘force multipliers’, and from a P&LEA perspective, new modus operandi (MO), tactics, techniques and procedures (TTPs) are discovered by bad actors to turn the capabilities of AI technology into enablers for crime, criminality and security threats.

The risks of using generative AI to create ‘deepfakes’ is already well known. Generative image models such as Midjourney (https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F) and Stable Diffusion (https://stability.ai/blog/stable-diffusion-public-release) continuously become more capable

[32] ‘Neo-Nazis and White Supremacists worldwide look to Artificial Intelligence (AI)’ Steven Stalinski et al; Middle East Media Research Institute (MEMRI), May 19 2023 (Accessed June 6 2023). https://www.memri.org/dttm/neo-nazis-and-white-supremacists-worldwide-look-artificial-intelligence-ai-—-national-security

[33] [Youtube video (2023, August 14). *How Cyber Criminals Are Using ChatGPT (w/ Sergey Shykevich)* [Video]. Yannic Kilcher. https://www.youtube.com/watch?v=10nEx2-8J0M

and the latter has also been recently released to the public. Generative video models are not as capable yet, but the capability for their misuse should not be underestimated.

Speech recognition models such as OpenAI:s Whisper (https://openai.com/research/whisper), and text to speech models also have the potential for misuse. There has already been a proliferation of effort, particularly from smaller organizations, to develop models for release to the public.

The usage of released models, either singly or in a combined manner, could perhaps provide further unforeseen pathways into misuse. A number of examples already exist that show the potential of using a modality where tools already exist as a method to learn a new modality where other advantages exist e.g. in the case of 'pose estimation' where using WiFi signals instead of cameras allows for pose estimation in the absence of light (https://youtu.be/xoVJKj8lcNQ?t=1117).

To sum up, it is impossible to know how and where this new wave of AI technology will impact on society. Consequently, P&LEAs must develop new competencies to understand it and be vigilant as to its effects and potentialities, while at the same time being aware of how they can utilise it in the service of society.

3.3 Some implications for the ALIGNER policy recommendations

In the rapidly evolving landscape of AI technology, Large Language Models (LLMs) and chatbots have emerged as transformative tools, reshaping the way we interact with technology and information. These advancements, while promising, also present challenges, particularly in the realms of accuracy, biases, and potential misuse. Once LLMs became potent, which only happened very recently, researchers and others have been caught by surprise with their emergent properties and the applicability of them. Thus, it is still impossible to know how they will impact society.

The proliferation of LLMs is two sided in terms of their benevolence for society. One may argue[34] that LLMs in the wrong hands impose serious security concerns. However, the monopoly of Microsoft, Alphabet and Meta is now broken, and alternatives have immediately been innovative in uses and improvements in terms of effectiveness and performance. Such innovation may also promote security solutions.

For policing and law enforcement agencies, the implications are multifaceted. LLMs can enhance operational efficiency, but they also open avenues for actions with malicious intent, from cyber-attacks to identity theft. As we navigate this technological frontier, stakeholders from developers to policymakers face the task of balancing the benefits and risks associated with these AI tools. It would be advisable for organizations involved in security, policing and law enforcement to be vigilant and to build competence about the new world of AI.

Over the last twelve months, both the slow pace of the on-going evolution of the EU AI Act proposal in the political arena and fast pace of new developments in AI technology and its public use have had effects in the environment within which P&LEAs operate. In combination, these two phenomena have made the environment both more uncertain for them to operate in and for its scope to become much wider.

[34] Schneier, B. (2023, June 2). Open-Source LLMs. Schneier on Security. https://www.schneier.com/tag/artificial-intelligence/page/3/

As a consequence, this also has implications for the ALIGNER policy recommendations and on ALIGNER's formal remit. AI technology is constantly and rapidly providing new threats, challenges and opportunities for PLEAs, yet it is almost impossible for ALIGNER to give meaningful policy recommendations "*tailored to the operational needs of the law enforcement sector"* when it is not yet known what P&LEAs are lawfully allowed to do, or are forbidden by law from doing, under the proposed EU AI Act.

4. Examining the ALIGNER policy recommendations in the wider policy environment

The opportunity has been taken by ALIGNER during its second year to try and assess where its original policy recommendations sit in the wider AI policy landscape. In order to do this, it was decided to assess how congruent each one of them might be when compared with the recommendations produced to date by ALIGNER's 'sister' projects in the EC AI and law enforcement cluster. There are two of these Projects; 'PopAI' (*European Positive Approach towards Artificial Intelligence tools in support of Law Enforcement & Safeguarding Privacy & Fundamental Human Rights*) and Project STARLIGHT (*Enhancing the EU's strategic autonomy in field of Artificial Intelligence for LEAs*).

Like ALIGNER, each of the above projects commenced work in 2021 and have produced their own policy recommendations.[35] PopAI was completed in September 2023 and hence their policy recommendations are now finalised, while STARLIGHT has recently (September) produced them in their first interim version. A final element in the comparison was added by selecting relevant recommendations and findings from a set of current recommendations produced by the European Union Agency for Cybersecurity (ENISA). ENISA released its recommendations in June 2023.

It is important to note that the comparative assessment was not carried out on an exact 'word for word' basis. This approach would not have been effective or useful due the myriad interpretations and differing contexts of many of the key words used by the projects e.g. "framework", "evaluation" "performance" or "standards". To mitigate this, the comparison was made by using a combination of the six ALIGNER policy recommendations from September 2022 and, for each one, a more broadly worded 'Aim' specifically written for each one.

4.1 Assessing the congruencies of the ALIGNER policy recommendations with other AI research cluster projects

An initial process of selection was undertaken to ensure that the other project policy recommendations used in the comparison were broadly compatible with the focus of the ALIGNER policy recommendations even if they were not specifically *"…tailored to the operational needs of the law enforcement sector…".* A total of twenty-eight selected recommendations were then assigned to what was judged to be the most relevant ALIGNER policy recommendation.

In addition, there were also four instances where policy recommendations from PopAI and STARLIGHT were considered as an appropriate 'fit' with two ALIGNER policy recommendations, and a further five when they were congruent with three ALIGNER policy recommendations. Overall, this meant that the six ALIGNER policy recommendations were compared with forty-two external policy recommendations from three different sources.

The detailed comparisons are shown below in Tables 3 through to Table 8 below and are followed by a numeric breakdown of the assignments as shown in Table 9.

[35] A summary of all three sets of policy recommendations can be found at Annexes A.1 to A.3 of this document.

Table 3 – Areas of Congruence: ALIGNER Policy Recommendation 1 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 1
Ensure the procurement, utilisation and in-service development of all AI technology by P&LEAs is carried out in a holistic way, with full cognizance of the adverse and beneficial impacts it may have on society *Aim of ALIGNER PR 1: To ensure the procurement, utilisation and in-service development of all AI technology adopted or used by P&LEAs is achieved in an integrated and consistent way*
Analogous policy recommendations/findings from other selected projects
popAI *(Policy Recommendations, September 2023)*
PR2: “EU to establish a common, harmonised European AI Regulation for LEAs that will govern the entire process from the design to the implementation and final use of AI systems by law enforcement authorities” **PR5**: “Establishment and standardization of a holistic impact assessment process” **PR6**: “Establishment of lawfulness, transparency and accountability protocols for LEAs” **PR7**: “EU to support the development of guidelines designed especially for the use of AI systems by LEAs” **PR8**: “Formulating a general ethical framework for the use of AI tools, taking into account that the AI Regulation can only regulate the basic legal framework for the use of AI” **PR10**: “To establish a procedure for the evaluation of AI tools by LEAs, from the ethical point of view” **PR11**: “EU Member States to support the continuous monitoring of AI systems in use, taking into account the perspective of civil society organisations” **PR14**: “Institutionalisation of multidisciplinary collaboration” **PR15**: “Establish inclusive AI development standards”
STARLIGHT *(WP4 Interim Policy Recommendations, September 2023)*
IPR2 “Coherent & consistent regulations – Differences in each legal system can become a barrier when developing AL tools in a multinational consortium and employing them afterwards” **IPR3** “Coherent and consistent ethical approaches - Ethical principles should be developed in a manner to allow different parties to apply them consistently, homogenously; clear and practical guidance should be developed into the tools, based on a common European understanding” **IPR5** “Contextual feasibility and coherence for LEAs…The tools/approaches that will be developed need to consider differences of country, language, operational environment and software infrastructure and [must] work to ensure inter-operability”

ENISA	*(‘Research Gaps’, June 2023)*
RG1 “Development of standardised data sets following these requirements in order to reliably reproduce and compare existing data sets” **RG2** “The need for a standardised performance evaluation framework to enable reliable comparisons between solutions addressing the same or similar problems”	

Table 4 – Areas of Congruence: ALIGNER Policy Recommendation 2 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 2	
Ensure there is always a competent and knowledgeable ‘human in the loop’ if AI technology is used to support and assist P&LEAs and criminal justice systems in critical decision-making processes ***Aim of ALIGNER PR 2***: *To ensure P&LEAs always have a trained, competent and knowledgeable ‘human in the loop’ when **critical decision-making processes** utilise AI technology*	
Analogous policy recommendations/findings from other selected projects	
popAI	*(Policy Recommendations, September 2023)*
PR9: “EU to ensure that the use of AI tools by police officers must be subject to multi-level control”	
STARLIGHT	*(WP4 Interim Policy Recommendations, Sept. 2023)*
None identified	
ENISA	*(‘Research Gaps’, June 2023)*
RG3 “Bringing ‘humans into the loop’ e.g. training practitioners using real-world scenarios”	

Table 5 – Areas of Congruence: ALIGNER Policy Recommendation 3 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 3
Review the EU data protection framework on the use by AI technology of ‘real-world’ data as it appears to act as a barrier to the procurement and use by P&LEAs of AI technology ***Aim of ALIGNER PR 3***: *To encourage a review of the EU data protection framework as it relates to the nexus of ‘real-world’ data, P&LEA functions and operations and the potential use of AI technology to access and process it as currently it appears to act as a barrier to their procurement and use of AI technology*
Analogous policy recommendations/findings from other selected projects

popAI	*(Policy Recommendations, September 2023)*
PR3: "EU to develop clear guidelines and standards for the collection, storage, restriction and use by LEAs of sensitive data, including biometric or other data" **PR17**: "EU to empower people to lodge a complaint and seek redress when their rights have been violated by the use of an AI system for law enforcement"	
STARLIGHT	*(WP4 interim Policy Recommendations, September 2023)*
IPR1 "Foreseeable regulations and guidance concerning data protection in scientific research: need for greater legal clarity, and a pressing need for more guidance and legislation on scientific research using sensitive data in order to enable AI innovation" **IPR2** "Coherent & consistent regulations – Differences in each legal system can become a barrier when developing AL tools in a multinational consortium and employing them afterwards"	
ENISA	*('Research Needs', June 2023)*
RN3 "Development of 34armonized34d frameworks assessing the preservation of privacy and the confidentiality of information flows as well as the designed system" **RN4** "Development of AI training models for practitioners using real-world scenarios"	

Table 6 – Areas of Congruence: ALIGNER Policy Recommendation 4 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 4	
Enable and support the P&LEAs of EU Member States to bridge gaps between their current levels of technology, the AI technology of today and AI technology already on the near horizon ***Aim of ALIGNER PR 4:*** *To enable and support the P&LEAs of Member States to bridge the gaps between their own current technology, the AI technology already available and the AI technology visible on the near horizon*	
Analogous policy recommendations/findings from other selected projects	
popAI	*(Policy Recommendations, September 2023)*
PR1: "EU to provide a legal framework for continuous AI training and educational programmes under the AI literacy notion for LEAs and civil society" **PR2**: "EU to establish a common, harmonised European AI Regulation for LEAs that will govern the entire process from the design to the implementation and final use of AI systems by law enforcement authorities" **PR7**: "EU to support the development of guidelines designed especially for the use of AI systems by LEAs" **PR8**: "Formulating a general ethical framework for the use of AI tools, taking into account that the AI Regulation can only regulate the basic legal framework for the use of AI"	

PR12: "Establish a European AI Systems Registry that will hold basic information about each AI system used by each LEA, by country, and its records to be accessible to all EU citizens"
STARLIGHT *(WP4, Interim Policy Recommendations, September 2023)*
IPR4 "Regulatory sandboxes – collaboration between technical, legal, ethical experts [and P&LEA users] in a dynamic setting" **IPR3** "Coherent and consistent ethical approaches – Ethical principles should be developed in a manner to allow different parties to apply them consistently, homogenously; clear and practical guidance should be developed into the tools, based on a common European understanding"
ENISA *('Research Needs', June 2023)*
RN1 "Test beds to study and 35armoniz the performance of ML-based tools and technologies used for cybersecurity"

Table 7 – Areas of Congruence: ALIGNER Policy Recommendation 5 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 5
The EU to approach AI technology adoption in the context of P&LEAs and criminal justice systems by using a Directive, as used to counter terrorism from 2017 onwards ***Aim of ALIGNER PR 5:*** *To encourage the EU to provide a tailored legislative framework designed specifically to guide and support Member State P&LEAs in their adoption and use of AI technology*
Analogous policy recommendations/findings from other selected projects
popAI *(Policy recommendations, September 2023)*
PR1: "EU to provide a legal framework for continuous AI training and educational programmes under the AI literacy notion for LEAs and civil society" **PR2**: "EU to establish a common, 35armonized European AI Regulation for LEAs that will govern the entire process from the design to the implementation and final use of AI systems by law enforcement authorities" **PR6**: "Establishment of lawfulness, transparency and accountability protocols for LEAs" **PR7**: "EU to support the development of guidelines designed especially for the use of AI systems by LEAs" **PR8**: "Formulating a general ethical framework for the use of AI tools, taking into account that the AI Regulation can only regulate the basic legal framework for the use of AI" **PR10**: "To establish a procedure for the evaluation of AI tools by LEAs, from the ethical point of view"

STARLIGHT *(WP4 interim policy recommendations, September 2023)*
IPR2 “Coherent & consistent regulations – Differences in each legal system can become a barrier when developing AL tools in a multinational consortium and employing them afterwards” **IPR3** “Coherent and consistent ethical approaches – Ethical principles should be developed in a manner to allow different parties to apply them consistently, homogenously; clear and practical guidance should be developed into the tools, based on a common European understanding” **IPR5** “Contextual feasibility and coherence for LEAs…The tools/approaches that will be developed need to consider differences of country, language, operational environment and software infrastructure and must work to ensure inter-operability”
ENISA *(‘Research Needs’, June 2023)*
RN5: “Establishing an observatory for AI and cybersecurity threats”

Table 8 – Areas of Congruence: ALIGNER Policy Recommendation 6 compared to selected policy recommendations from external projects

ALIGNER Policy Recommendation 6
Conduct further and regular research into P&LEA and criminal justice concerns and capability needs for AI technology in order to ensure EU policy makers are aware of new developments and on-going issues ***Aim of ALIGNER PR 6***: *To urge the EU to conduct further, regular research into AI technology solutions as a means to assist P&LEAs to identify ‘real-world’ problems, to define their operational needs and to determine where capability gaps exist*
Analogous policy recommendations/findings from other selected projects
popAI *(Policy Recommendations, September 2023)*
R13: “EU funding investment for research and development in order to explore the use of AI systems in LEAs”
STARLIGHT *(WP4, Interim Policy Recommendation, September 2023)*
None identified
ENISA *(‘Research Needs’, June 2023)*
RN2: “Development of penetration testing tools based on AI and ML to find and exploit security vulnerabilities to assess the behaviour of attackers”

Once the external policy recommendations had been collated, assessed and assigned to one or more ALIGNER policy recommendations, it became possible to see in numeric terms where their congruencies lay (see Table 9).

Table 9 – External Project comparisons: Number of congruent Policy Recommendations from external projects assigned to each ALIGNER policy recommendation (n=42)

Project ALIGNER Policy Recommendation Nos. (PR)	Total Project popAI PRs congruent with ALIGNER	Total Project STARLIGHT PRs congruent with ALIGNER	Total ENISA PRs congruent with ALIGNER
ALIGNER PR 1	**9** (PRs.2,5,6,7,8,10,11,14,15)	**3** (PRs.2,3,5)	**2** (RG1, RG2)
ALIGNER PR 2	**1** (PR.9)	**Zero**	**1** (RG3)
ALIGNER PR 3	**2** (PRs.3,17)	**2** (PRs.1,2)	**2** (RN3, RN4)
ALIGNER PR 4	**5** (PRs.1,2,7,8,12)	**2** (PRs.3,4)	**1** (RN1)
ALIGNER PR 5	**6** (PRs.1,2,6,7,8,10)	**3** (PRs.2,3,5)	**1** (RN5)
ALIGNER PR 6	**1** (PR.13)	**Zero**	**1** (RN2)
Total	**24**	**10**	**8**

A number of insights can be drawn from these results and are set out in the following section.

4.1.1 Findings from the cross-project congruency assessment

On examination from a high level cross-project perspective, the congruency assessment yielded a number of insights into the overall policy recommendation environment of them all, including those of ALIGNER. In essence, these insights are predominantly associated with the need to provide certainty for P&LEAs relating to how they can acquire and utilise trusted AI technology in a holistic way, and how they can do so within a regulatory environment that both supports and facilitates P&LEA operations and functions. A second, smaller category of suggestions refer to specific requirements to enhance the functions and activities of P&LEAs.

The most important policy recommendations were aggregated and then summarised as follows:

- The procurement, utilisation, in-service development and use of AI technology tools by P&LEAs or on their behalf should be approached in a holistic way. This should be based on a common European understanding of AI technology and carried by P&LEAs utilising common and harmonised EU legal directions and guidelines

- For P&LEAs to effectively engage with AI technology and its potential capabilities, there is a need for an EU-wide framework to enable, enhance and institutionalise how they do so. It should be designed for this purpose and utilise elements of primary legislation, statutory regulations, guidelines and protocols, systems and structures, processes and procedures

- The role of ethics in the P&LEA approach to AI technology and tools is seen to be of fundamental importance and should be governed by a clear and consistent ‘ethical framework’ of principles and practices

- AI technology tools for P&LEAs in the EU should be legally compatible and functionally interoperable. To achieve this, they should be evaluated (for legality, ethical compliance and functional performance). Consequently, there is a need for agreed standards to enable them to be assessed and compared

- The establishment by the EU of a single body or organisation to act as a nexus between rapidly developing AI technologies and their impacts on P&LEA functions and responsibilities could bring many benefits. Two of its suggested functions are as an Observatory to monitor AI and cybersecurity threats (ENISA) and as a central AI Systems Registry of each one used by P&LEAS across the EU (STARLIGHT)

  Other functions could also be envisaged for it, for example as a source of impartial and accurate advice and guidance for P&LEAs on issues relating to AI technology. This type of approach would contribute to a common European understanding and enhance P&LEA cooperation and collaboration

It should be noted that if the policy recommendation insights are embraced as a whole, then the more likely they are to enhance the effectiveness of P&LEAs whilst at the same time safeguarding the rights and privacy of individuals and wider society. An alternative future dominated by unguided and unregulated deployment of AI technology by P&LEAs is highly unlikely to achieve any comparable situation.

Finally, the congruency assessment can also be used as a benchmark against which the ALIGNER policy recommendations can be re-evaluated and revised as necessary, as described below in Section 4.1.2.

4.1.2 ALIGNER Policy Recommendation revisions in the light of the congruency assessment

The original set of ALIGNER policy recommendations published in 2023 were deliberately not ranked into any order of priority as it was felt that it was too early in the project for this to be determined in any meaningful way. Consequently, their numbering from 1 to 6 was done only to distinguish between them for reference purposes.

After the congruency assessment process, it became possible to develop this by ranking each ALIGNER policy recommendation according to its degree of congruence with the external projects i.e. from the most congruent (with 14 external project policy recommendations) to the least congruent (with 2 external project policy recommendations). Once this was done, the ALIGNER policy recommendations were reordered and renumbered. This step is shown in detail in Table 10 below:

Table 10 - ALIGNER Policy Recommendations ranked in descending order of congruency compared to those of selected external projects

ALIGNER PR ranked in order of congruency	ALIGNER Policy Recommendation Nos.	popAI PRs congruent with each ALIGNER PRs	STARLIGHT PRs congruent with each ALIGNER PRs	ENISA PRs congruent with each ALIGNER PRs	Total no. PRs congruent with each ALIGNER PR
First	ALIGNER PR 1	9	3	2	(14)
Second	ALIGNER PR 5	6	3	1	(10)
Third	ALIGNER PR 4	5	2	1	(8)
Fourth	ALIGNER PR 3	2	2	2	(6)
Fifth	ALIGNER PR 2	1	0	1	(2)
Sixth	ALIGNER PR 6	1	0	1	(2) **[Total = 42]**

Once they were ranked in order of concordance with the policy recommendations of the selected external projects, the opportunity was then taken to revisit and revise the wording of the original ALIGNER policy recommendations in order to better reflect the overall aim of each one. The new versions of these are shown in Table 11 below.

Table 11 – ALIGNER policy recommendations, 2023, with updated wording and ranked in order of congruency with selected external projects

ALIGNER Policy Recommendation 1 (Revised 2023)
The EU to ensure that the procurement, utilisation and in-service development of all AI technology, by or on behalf of P&LEAs, is carried out in a holistic and consistent way, taking full cognizance of both the adverse and beneficial impacts it may have on society
ALIGNER Policy Recommendation 2 (Revised 2023)
The EU to provide as a matter of priority a tailored legislative framework designed specifically to guide and support the P&LEAs of Member States in their adoption, acquisition and use of AI technology
ALIGNER Policy Recommendation 3 (Revised 2023)
The EU to assist and support the P&LEAs of Member States to bridge the gaps between their current level of technology and the AI technology that is already

available and to transition towards the AI technology that is already foreseeable on the near horizon
ALIGNER Policy Recommendation 4 (Revised 2023)
The EU to instigate a review of the EU data protection framework as it relates to the nexus between 'real-world' data and P&LEA functions and operations, and their potential use of AI technology to access and process it. Currently, it appears to act as a barrier to P&LEA procurement and use of AI technology
ALIGNER Policy Recommendation 5 (Revised 2023)
The EU and Member States to ensure P&LEAs always have a trained, competent and knowledgeable 'human in the loop' when they utilise AI technology to assist in decision making processes
ALIGNER Policy Recommendation 6 (Revised 2023)
To urge the EU to conduct and facilitate systematic and regular research into AI technology that may have potential ramifications (both as solutions and as threats) for the operational needs and capability requirements of P&LEAs

The wording and overall relevance of the ALIGNER policy recommendations will continue to be reviewed until they are submitted in a final version at the end of the project in September 2024. For example, it is envisaged that ALIGNER PR 6 (focused on research involving AI technology with potential ramifications for P&LEAs), will likely need to be revised during 2024 as a clearer understanding is arrived at of what are the most relevant aspects/topics of research from an ALIGNER perspective. Consequently, its wording is likely to change and this in turn may result in better congruence with the current research focused recommendations from the other projects.

In addition, information gathered between May and September 2023 from sixty-six respondents to the second ALIGNER on-line survey of practitioners from policing and law enforcement, as well as experts engaged in AI technology research and policy making, will be fully analysed and used to inform the policy recommendations as necessary.[36]

One early finding from this survey is that while AI technology is being used today by P&LEAs, it is to a very limited extent and involves a very limited number of organisations. However, the focus of its use seems to be for both biometric recognition (including video and photo analysis), as well as data and information handling processes. This underscores the importance of the ALIGNER policy recommendations (and other projects) highlighting the need for frameworks to encompass the identified ethical, legal and assessment methodology aspects (See Table 11 above, particularly ALIGNER policy recommendations 1, 2 and 4).

[36] The first ALIGNER survey (May to August 2022, 52 responses) and its results were published previously. See Daniel Lückerath, Valerie Wischott, Donatella Casaburo, Lindsay Clutterbuck, Peter Svenmarck, Tommy Westmann: ALIGNER D5.5 Research Roadmap for AI in Support of Law Enforcement and Policing – Version 2; H2020 ALIGNER, GA no. 101020574

5. Conclusion and next steps

During the single year that has passed since the ALIGNER policy recommendations were published, the development and public use of AI technology has moved forward extremely rapidly and from a P&LEA perspective, not all of it has been beneficial. The new capabilities and public availability of on-line chatbots based on algorithm driven Large Language Models has increased the challenges they face by a significant degree.

On the other hand, the Artificial Intelligence Act Proposal, the keystone EU legislation designed to encompass how they can respond to criminal and security threats and in addition, how and when they can utilise AI technology to do so, has not yet become EU law. While it is a taking the normal path for EU-wide legislation, with three-way 'trilogues' underway between the Commission, the Council and the Parliament to debate and agree its final provisions, P&LEAs have no certainty nor even clarity as to the outcome of some critical issues. For example, will they be allowed to use AI driven facial recognition technology and how will it be regulated, or will they be banned from using it under any circumstances? Will AI systems used for big data crime analytics be removed from the current category of 'High Risk" and if so, will they then be considered as 'Low Risk'? Both of these issues and many others directly impacting on P&LEAs appear to be still awaiting debate and resolution.

Against this background, ALIGNER has continued to develop its original policy recommendations by subjecting them to a cross project congruency assessment involving its two 'sister' projects of 'PopAI' and 'STARLIGHT' and also including recent recommendations from ENISA. This has shown that ALIGNER has a high degree of concordance with them across a number of policy recommendations and particularly coalesces with them where they propose suggestions and recommendations focused on how P&LEAs can utilise AI technology in a holistic way and how an operating environment can be created to both support and regulate their actions.

As a consequence of the cross project congruency assessment, the ALIGNER policy recommendations have been re-ranked to emphasise their wider concordance and their wording has also been revised to ensure that they continue to be *"…tailored to the operational needs of the law enforcement sector…".* With these revisions in place, consideration can be given to the next steps as ALIGNER enters into its third and final year.

Consideration will be given to using the initial results as a foundation from which to repeat the cross congruency assessment but this time, the comparison could be drawn from a greater number of relevant recommendations, as derived by a wider range of projects that have remits to engage with the topic of AI related technology and its interactions with P&LEAs. A final decision will be made after a feasibility and scoping exercise has been carried out.

Finally, by working collaboratively with Project STARLIGHT, ALIGNER will continue to examine both emerging AI tech developments and the progress of the AI Act as affects P&LEAs. It will also ensure that the policy recommendations are also informed by other initiatives as they develop. These include the standardisation efforts initiated in May 2023 by the EU when the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC) were formally tasked "to draft new [standards]…in support of Union policy on artificial intelligence."[37] and any new efforts launched under the European Security Data Space for Innovation "to increase trust in the use of Artificial Intelligence by law enforcement."[38]

[37] See https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en

[38] See https://home-affairs.ec.europa.eu/news/call-proposals-funds-data-sets-european-data-space-innovation-2023-03-30_en)

References

Casaburo, D., and Marsh, I. (2023). ALIGNER D4.2 Methods and guidelines for ethical & law assessment, available at https://aligner-h2020.eu/wp-content/uploads/ALIGNER-D4.2-Methods-and-guidelines-for-ethical-law-assessment-v20230324-FINAL.pdf (accessed on 22 May 2023).

Council of the European Union (2022). 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 - General approach, available at https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf (accessed on 21 August 2023).

Eren, E., Casaburo, D., and Vogiatzoglou, P. (2021). ALIGNER D4.1 State-of-the-art reports on ethics & law aspects in Law Enforcement and Artificial Intelligence, available at https://aligner-h2020.eu/wp-content/uploads/ALIGNER-D4.1-SotA-reports-on-ethics-law-aspects-v20220714.pdf (accessed on 17 August 2023).

European Commission (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (accessed on 18 August 2023).

European Commission (2020). White Paper on Artificial Intelligence – A European approach to excellence and trust, available at https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed on 18 August 2023).

European Parliament (2023). Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, available at https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf (accessed on 22 August 2023).

High-Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI, available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 (accessed on 18 August 2020)

Lückerath, D., Wischott, V., Casaburo, D., Clutterbuck, L., Svenmarck, P., Westmann, T., ALIGNER D5.5 Research Roadmap for AI in Support of Law Enforcement and Policing – Version 2; H2020 ALIGNER, GA no. 101020574

OECD Council (2019). Recommendation of the Council on Artificial Intelligence, available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed on 22 August 2023).

Ursula von der Leyen (2019). A Union that strives for more: My agenda for Europe, available at https://commission.europa.eu/system/files/2020-04/political-guidelines-next-commission_en_0.pdf (accessed on 18 August 2023).

Annexes

Annex A.1 – popAI Policy Recommendations (Second Policy Brief, Deliverable D1.7, September 30th 2023)

Policy Recommendation 1

"EU to provide a **legal framework for continuous AI training and educational programmes** under the AI Literacy notion for LEAs and civil society"

Policy Recommendation 2

"EU to establish **a common, harmonized European AI regulation for LEAs** that will govern the entire process from the design to the implementation and final use of AI systems by law enforcement authorities"

Policy Recommendation 3

"EU to develop clear **guidelines and standards for the collection, storage, restriction and use by LEAs of sensitive data, including biometric or other data**"

Policy Recommendation 4

"Carrying out **impact assessments** even in cases when this is not obligatory by law"

[NB - No close concordance assessed with ALIGNER PRs]

Policy Recommendation 5

"Establishment and standardization of a **holistic impact assessment process**"

Policy Recommendation 6

"Establishment of **lawfulness, transparency and accountability protocols for LEAs**"

Policy Recommendation 7

"EU to support the development of **guidelines designed especially for the use of AI systems by LEAs**"

Policy Recommendation 8

"Formulating a **general ethical framework for the use of AI tools**, taking into account that the AI Regulation can only regulate the basic legal framework for the use of AI"

Policy Recommendation 9

"EU to ensure that the **use of AI tools by police officers must be subject to multi-level control**"

Policy Recommendation 10

“To establish a procedure for the **evaluation of AI tools by LEAs**, from the ethical point of view”

Policy Recommendation 11

“EU Member States to support the **continuous monitoring of AI systems** in use, taking into account the perspective of civil society organisations”

Policy Recommendation 12

“Establish a European **AI Systems Registry** that will hold basic information about each AI system used by each LEA, by country and its records will be accessible to all EU citizens”

Policy Recommendation 13

“**EU funding investment for research and development** in order to explore the use of AI systems in LEAs”

Policy Recommendation 14

“Institutionalisation of **multidisciplinary collaboration**”

Policy Recommendation 15

“Establish **inclusive AI development standards**”

Policy Recommendation 16

“**Inclusion of Civil Society**”

[NB - No close concordance found with ALIGNER PRs)

Policy Recommendation 17

“EU to **empower people to lodge a complaint and seek redress** when their rights have been violated by the use of an AI system for law enforcement”

Annex A.2 - STARLIGHT Interim Ethical and Legal focused Policy Recommendations (As of September 2023)

Policy recommendation 1

"Foreseeable regulations and guidance concerning data protection in scientific research: need for greater legal clarity, and a pressing need for more guidance and legislation on scientific research using sensitive data in order to enable AI innovation"

Policy recommendation 2

"Coherent & consistent regulations – Differences in each legal system can become a barrier when developing AL tools in a multinational consortium and employing them afterwards"

Policy recommendation 3

"Coherent and consistent ethical approaches - Ethical principles should be developed in a manner to allow different parties to apply them consistently, homogenously; clear and practical guidance should be developed into the tools, based on a common European understanding"

Policy recommendation 4

"Regulatory sandboxes – collaboration between technical, legal, ethical experts [and P&LEA users] in a dynamic setting"

Policy recommendation 5

"Contextual feasibility and coherence for LEAs…The tools/approaches that will be developed need to consider differences of country, language, operational environment and software infrastructure and must work to ensure inter-operability"

Annex A.3 – Selected ENISA Research and Innovation Recommendations in concordance with ALIGNER Policy Recommendations (ENISA Research and Innovation Brief, Published June 7th 2023)

Research Gaps

Research Gap 1

"Development of standardised data sets following these requirements in order to reliably reproduce and compare existing data sets"

Research Gap 2

"The need for a standardised performance evaluation framework to enable reliable comparisons between solutions addressing the same or similar problems"

Research Gap 3

"Bringing 'humans into the loop' e.g. training practitioners using real-world scenarios"

Research Needs

Research Need 1

"Test beds to study and optimise the performance of ML-based tools and technologies used for cybersecurity"

Research Need 2

"Development of penetration testing tools based on AI and ML to find and exploit security vulnerabilities to assess the behaviour of attackers"

Research Need 3

"Development of standardised frameworks assessing the preservation of privacy and the confidentiality of information flows as well as the designed system"

Research Need 4

"Development of AI training models for practitioners using real-world scenarios"

Research Need 5

"Establishing an observatory for AI and cybersecurity threats"