CEASEFIRE

Ceasefire: Advanced versatile artificial intelligence technologies and interconnected cross-sectoral fully-operational national focal points for combating illicit firearms trafficking

D1.5 – Applicable legal/ethical framework and social impact assessment. Intermediate version

This project has received funding from the European's Union Horizon Europe research innovation programme under Grant Agreement No.101073876

Authors

Name	Partner	e-mail
Francesca Trevisan	P10 TRI	Francesca.trevisan@trilateralresearch.com
Christine Andreeva	P10 TRI	Christine.andreeva@trilateralresearch.com

Executive Summary

This deliverable provides an overview of how CEASEFIRE partners are embedding the ethics principles of trustworthy AI in the technology development up to M18. The deliverable also provides an overview of the legal requirements that the tools must abide by when -and if- reaching the use phase. Based on the ethical assessment, this deliverable provides a compilation of prerequisites for the training material that is being developed by CEASEFIRE partners. Finally, it presents a mid-term iteration of the societal impact assessment.

.

Table of contents

List of tables

List of acronyms

Acronym	Explanation
AI	Artificial Intelligence
ALTAI	Assessment List for Trustworthy AI
GCE	Graph Correlation Engine
GDPR	General Data Protection Regulation
LEA	Law Enforcement Agencies
LED	Law Enforcement Directive

1. Introduction

Between 2016 and 2017 across 81 countries, a total of 550,000 firearms were seized [1]. Still, this data seems to underestimate the real figures. Firearm trafficking is a key part of organised crime group activities in Europe. Their exchange and availability increase the risk of their use in terrorist attacks and organised crime activities. In fact, according to EMPACT [2], tackling illicit firearms trafficking is one of the EU's priorities in fighting serious and organised crime.

In response to this need, the CEASEFIRE project is developing a set of technologies aimed at improving the operational capabilities of EU Law Enforcement Agencies in detecting, analysing, and tracking cross-border illicit firearms trafficking-related activities. CEASEFIRE technologies are being designed to be used to respond to firearms trafficking threats and incidents after they have occurred, rather than to prevent or detect them beforehand. To achieve this purpose, the project consists of five different use cases, each one designed to address distinct user needs.

Use case 1: "Real-time systematic firearms incident and intelligence information collection and exchange" would allow users to collect, extract and analyse near real-time information on firearm incidents from online news articles. Using the incident tracking tool, users would be able to create digital reports and share strategic intelligence among law enforcement agencies.

Use case 2: "On the spot firearm seizure registration and cross-border data search" would allow users to automatically identify key characteristics of firearms seized at a crime scene, look up information on the firearm in databases and register the seized firearms with comprehensive, accurate and detailed information on their characteristics. Users would be able to identify on site the firearm type, model and other critical information, as well as automatically search and retrieve information from relevant databases regarding the seized firearms.

Use case 3: "Firearms purchase on dark web marketplaces" would allow the collection, review and analysis of information from the dark web on illegal firearms marketplaces and forums, to gather information on online actors involved in illegal firearms trading and to identify possible links that could lead to the discovery of their identities.

Use case 4: "mail order and courier service firearms trafficking detection using scanning technologies" involves automatically detecting - through an x-ray parcel scanner - and identifying illegal firearms, critical components and ammunition within parcels sent via postal and courier services in the EU.

Lastly, use case 5: "3D printed firearm blueprints distribution" aims to gather, review, and analyse information from online sources, including forums and social media platforms, regarding blueprints for 3D-printed firearms. The objective is to collect information that could lead to the identification of online actors/entities who are sharing these blueprints and to uncover possible links leading to the identification of systematic distributors and their distribution networks.

This deliverable contains the information regarding how the consortium is putting into practice trustworthy AI development. In this regard, TRI assessed each technical task against the Assessment List for trustworthy AI (ALTAI) to identify how ethics principles are being embedded in the design of CEASEFIRE technology. As a result, this deliverable presents practical information on how considerations on human agency, technical robustness, privacy, transparency, fairness and non-discrimination, societal and environmental wellbeing, and accountability are being integrated in the project in the technologies. Following this evaluation, this deliverable outlines some key legal requirements based on the Law Enforcement Directive and the AI Act. Finally, it provides an updated version of the societal impact assessment.

It is important to note that most research activities are in progress and the information provided in this report might be subject to changes.

1.1. Version specific notes

This deliverable includes the mid-term ethics evaluation of the development activities related to CEASEFIRE technologies. The final version of the legal and ethical assessment is scheduled for completion in the project's final month, September 2025. While researchers carried out a detailed task by task ethics assessment, to preserve the security of the technologies that are being developed, this public version of the deliverable presents data in an aggregated manner and does not refer to specific tasks.

2. Responsible development from theory to practice

Below we outline how CEASEFIRE is embedding the 7 principles for trustworthy AI [3] during the technology development phase. The seven principles include human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being and accountability. Following these non-binding principles support the design of human-centric technologies that are aligned with the values outlined in the Charter of Fundamental Rights [3].

It is worth mentioning that as of M18 in the project timeline, CEASEFIRE partners have not completed the development of the technology layers and the majority of research activities are ongoing. Consequently, some of the approaches outlined in the sections below might undergo changes.

As part of this work, TRI distributed an ethics questionnaire and engaged in ongoing dialogues, conducting both bilateral and multilateral meetings with technical partners to discuss the functionalities, intended purposes, design decisions, and other pertinent aspects of the tools. Throughout these interactions, TRI has supported partners to embed ethics considerations in the design of the technology layers.

2.1. Human agency and oversight

Preserving human agency and oversight is the first principle to embed in trustworthy technology. AI systems should support, rather than replace, human decision-making, and users should be able to make informed autonomous decisions regarding AI systems [3]. To achieve this, users should be given the knowledge and tools to understand and interact with AI systems to a satisfactory degree and, where possible, be enabled to self-assess or challenge the system.

In CEASEFIRE, users will maintain meaningful control over the most important aspects of decision-making process through several measures.

First of all, users will receive training on how CEASEFIRE technologies operate, their advantages, and their limitations. Additionally, users will be instructed on how to understand and interpret the results or outputs produced by CEASEFIRE products. Training modules for these purposes are currently being developed throughout the project period. Beyond training, users will have available the technical documentation with information about:

- intended purpose
- the methods and steps performed for the development of CEASEFIRE products
- the general logic of CEASEFIRE products
- the key design choices including the rationale and assumptions
- classification choices
- description of the system architecture
- training methodologies
- validation and testing procedures used
- information about the monitoring, functioning and control of the AI system

Second, CEASEFIRE partners are developing appropriate human-machine interfaces that will enable the user to oversee the system during its use and better understand and interpret its outcome.

Third partners are implementing design decisions that facilitate users critical thinking on CEASEFIRE outcomes. For example, in use case 2 users will take a picture of weapon and the identification will show in order of likeness. Users then will have to choose what seems for them the most probable. If none seem probable, then they will be able to click on a button can say none is correct. This button is the option to override an identification. Additionally, to reconstruct the network of firearms trafficking, users will be provided with certainty levels of different options. Users will understand how to interpret certainty levels through the training program developed during the project duration. From this, end users will be able to make assumptions and build graphs based on network associations. End users will be therefore the last to judge whether the result is valid or not.

CEASEFIRE partners are also implementing explainability aspects through user friendly and easily interpretable visual analytics which will show how the task from which the node come. Partners are providing a framework for the visualization of firearms trafficking trails that will facilitate users to easily create different tailored dashboards, based on the case being examined every time.

In terms of preserving autonomy, it is important to note that CEASEFIRE products do not seem able to directly link the digital traces with the identity of a perpetrator. LEAs themselves will be the only persons able to link, directly or indirectly, these evidence to a natural person.

2.2. Technical robustness and safety

Trustworthy AI systems must be safeguarded against vulnerabilities that could be exploited by adversaries, such as hacking [3]. Additionally, AI systems should be accurate and reliable.

The table below summarises the key measures to embed robustness and safety in the CEASEFIRE technologies.

Table 1 Technical robustness and safety.

Measure	Description
Access control	Access to the CEASEFIRE system will be controlled with authentication, authorisation and accounting. Access control has been defined as a mandatory non-functional requirement for the most critical components and optional but highly recommended for the components of all other tasks.
Communication protocols	Partners are using secure, robust and industry standard communication protocols such as Kafka and secure AJAX REST APIs.
Logging	The system will provide logging of actions. All interactions with the system must be documented through the identification of the users and the specific activities they conducted with the system.
Evaluation of outputs	Outputs generated by the models undergo a rigorous evaluation and filtering process by researchers and LEA experts. This critical assessment ensures that the findings produced by the system are accurate, relevant, and reliable, aligning with expert knowledge and operational standards.
Data quality	Accuracy and reliability metrics will be reported. Monitoring, documentation, and optimization of the system's accuracy are continually performed, with a focus on prioritizing communication of these metrics to end users.
Vulnerability scanning	Partners a vulnerability scanner, across all the CI/CD pipelines they construct. The system operates within

	a closed circuit, mitigating exposure to cyberattacks, and it is fortified by certification mechanisms to ensure robust security measures.
Confidence levels shown to users	Users will be provided with confidence levels of outputs so they can assess the degree of uncertainty.
Robustness metrics	Partners are evaluating the use of robustness metrics to quantify the susceptibility of the firearms detector to adversarial attacks, as well as adversarial training to pre-emptively mitigate such vulnerabilities.

2.3. Privacy and data governance

AI systems should ensure privacy and data protection throughout their entire lifecycle. This section details the key privacy by design measures that partners are implementing during the project's duration.

Regarding the crawler, partners are conducting a **highly targeted** crawling activity based on a set of firearms-related key words (e.g. model, brand) produced as part of an ontology during the project. This approach narrows significantly reduces the risk of inadvertently collecting data that are not relevant to the scope of the project.

To ensure data security throughout transit and storage, partners are employing robust encryption protocols.

During the pre-processing phase, partners are using anonymization and pseudonymization techniques to protect personal data. Before processing data through the CEASEFIRE algorithms, partners are:

- removing identifiers such as any names, addresses, and other direct identifiers are stripped from the dataset.
- generalizing data, which involves converting specific details into broader categories to protect privacy. For instance, instead of using exact geolocation data, only higher-level location information is used.
- Aggregating data to show trends and patterns without revealing individual-level information.

Partners are developing training materials which will include guidelines for users on responsible use of the crawler during criminal investigations.

2.4. Transparency

The transparency requirement is linked to the principles of traceability, explainability and transparent communication to users. In other words, users should know they are interacting with an AI system, and they should be able to understand its inner processing and outputs [3].

CEASEFIRE research participants and end-users will be informed about:

(1) their interaction with an AI system/technology;

(2) the abilities, limitations, risks and benefits of the AI system/technique;

(3) the manner in which decisions are taken and the logic behind them.

1. **Interaction with the AI system**

- Research participants will be informed about their interaction with AI prior to the participation to any research activity and training. Information about how the AI system is working is being provided in an understandable language and a research partner is being always available for questions.
- End users will go through an in-depth training session before using the AI system. The training package is being developed as part of the project activities. The training material will clarify how users will interact with CEASEFIRE technologies.
- Partners are adding a disclaimer to the platform that reminds end users that the output they see has been produced by using AI.

2. **About the abilities, limitations, risks and benefits of the AI system/technique**

- Research participants will be informed about the abilities, limitations, risks and benefits of the AI system before as well as after participating to the research activity. When, during trainings and pilots, participants will interact with the AI systems partners will be always available to answer questions and will explain users each functionality with tailored content before any interaction taking place.
- End users will be informed about the abilities, limitations, risks and benefits of the AI system through the training package developed as part of WP8. The first version of the training material will be submitted in M22. The explainability feature will also make users available information on how the outputs are computed.

3. **About the manner in which decisions are taken and the logic behind them**

- CEASEFIRE platform and the functionalities within it are not supposed to take decisions but to support LEAs' decision making. Users will be informed about their interaction with AI through a disclaimer (e.g. "*This output has been produced by an AI model. Click here to know more*".) on the front end. Users will also be able to have more information on how the output was obtained (by clicking on a "*click here for more information*" button). The training material developed as part as WP8 will also contain in depth information on the logic behind each functionality of the platform. The first version of the training package will be delivered in M22.
- Research participants will be informed about the logic behind the technology that is being piloted before taking part to the pilot studies. During the pilots and trainings, technical partners will be available to answer questions.

To allow traceability and increase transparency, the data sets and the processes that yield the AI system's output, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard [3]. In this regard, CEASEFIRE partners are documenting all the development process, including the information on the data used and design choices in a number of deliverables that are regularly submitted to the European Commission.

2.5. Diversity, non-discrimination and fairness

Embedding diversity, non-discrimination and fairness throughout the AI system's life cycle is key to achieve trustworthy AI [3]. This means that AI systems should avoid unfair bias, be accessible and involve in the development process affected stakeholders. CEASEFIRE partners are taking measures to prevent, avoid and mitigate potential bias data, design choices and the usage of CEASEFIRE products.

a. **Data**

The table below provides a list of the techniques that partners are using to mitigate bias in the data.

Table 2 Techniques to mitigate bias.

Fairness metrics	Using metrics for demographic parity, equal opportunity, and predictive equality to regularly audit and assess the performance of algorithms across different groups
Adversarial debiasing	Training a model to predict the target variable accurately while minimising the ability to predict a sensitive attribute (e.g., race or gender) from the model's predictions.
Use of bias mitigation algorithms	Algorithms designed specifically to reduce bias at different stages of the ML process—pre-processing (modifying data to remove bias before training), in-processing (modifying the learning algorithm itself), and post-processing (adjusting the model's predictions). Techniques such as re-weighting training examples or applying fairness constraints fall into this category
Feature selection and engineering	Selecting and engineering features to exclude those that directly or indirectly encode biased or sensitive attributes. This might involve removing features that are closely correlated with sensitive attributes or creating new features that better represent individuals from all groups.
Regularisation techniques	Methods that penalise the complexity of the model can prevent overfitting to biased patterns in the training data. Techniques like L1 or L2 regularisation can be particularly effective in making models generalise better, reducing the chance of biased predictions.
Sensitive Attribute Blindness	Excluding sensitive attributes (e.g., race, gender) from the model training process.

b. Design choices

TRI is implementing a granular and iterative approach to support partner in decision-making and assess legal, ethical and social implications of CEASEFIRE design choices. TRI role is to support CEASEFIRE partners in carrying out the activities following a responsibility-by-design principle and in compliance with ethics and legal frameworks. In practice, adopting a responsibility-by-design approach means mapping out the legal, ethical and societal risks and mitigation measures iteratively, while involving different types of stakeholders and supporting partners to understand the provisions of relevant regulations (e.g., AI Act) and ethics principles (e.g. AI HLEG Ethics Guidelines on

Trustworthy AI). This supports partners to balance out different needs, legal and ethical requirements and take decisions about technology features and functionalities.

TRI is iteratively assessing the impact of CEASEFIRE activities on ethics, data protection, human rights (including discrimination related issues) using the Trilateral Touchpoint Table™ methodology and the Assessment List for Trustworthy AI. The Trilateral Touchpoint Table™ was used to map initial risks. The Assessment List for Trustworthy AI was used to understand how each task is embedding trustworthy AI principles in the design choices.

On another note, the development of CEASEFIRE technologies is an iterative process that actively involves consulting with future end users to obtain their regular feedback. This engagement aims at align CEASEFIRE technologies with their needs and expectations. Beyond end-users, CEASEFIRE is involving a wide range of stakeholders, including legal experts, ethicists, social psychologists, and policymakers in the development process to include diverse perspectives and address potential concerns. The final version of this deliverable, which is scheduled for completion by month 36 (September 2025), will also incorporate the perceptions and expectations of citizens.

c. Use

Discrimination and bias at use phase is being prevented through:

- Training: CEASEFIRE is embedding ethical considerations on bias and discrimination in the training modules that are being developed by partners. For example, the first training included information on the Ethics Guidelines for Trustworthy AI and involved users on brainstorming how diversity and non-discrimination are relevant to CEASEFIRE systems. Each training session is presenting an opportunity to assess the adequacy of the training materials and make necessary adjustments to better align the final training content with the needs of end-users.
- Recommendation for periodic audits: partners suggested end-users periodic audits to monitor and uncover potential biases at use case. This recommendation will be included in the training material.
- Explainability features: partners are incorporating in CEASEFIRE some explainability features that will help users understanding how the final output has been produced and obtaining more information on how the outcome has been computed.

2.6. Societal and environmental wellbeing

To be deemed trustworthy, AI systems should incorporate considerations on environmental and societal well-being [3]. This involves designing and implementing AI solutions that not only minimize negative environmental impacts, but also actively contribute to the betterment of society.

Recognising the environmental implications of CEASEFIRE technological solutions, CEASEFIRE partners have optimised the GCE for computational efficiency and adopted green computing practices to minimise its ecological footprint. Additionally, the system's design for adaptability ensures it remains effective against the evolving dynamics of criminal activities, contributing to long-term societal benefits.

As a decision support component, the CEASEFIRE system is carrying out the type of analysis that investigators currently do manually. However, there are not enough investigators with the skills to perform these task, which is not expected to change in the future. The component frees time for investigators to work on non-repetitive tasks and to focus on the “understanding” of the rationale behind observed firearms related criminal activities and on the analysis tasks of higher complexity. Criminal activities and modus operandi change fast, human subject-matter experts will remain necessary to update the system. Appropriate understanding of the results will require some conceptual level understanding of statistics and probability which may not currently be present. As an outcome of the project, training and re-skilling programs will be offered to the workforce.

2.7. Accountability

The accountability requirement is intrinsically tied to the principle of fairness. It demands the establishment of mechanisms to ensure responsibility and accountability for AI systems and their outcomes at every stage—prior to, during, and following their development, deployment, and use [3]. CEASEFIRE is advancing accountability through various means.

First of all, CEASEFIRE consortium is writing and submitting to the European Commission detailed documentation (deliverables) of the system's development processes, including design decisions, algorithm choices, data sources, and model training procedures.

Second, partners are working to implement strict access controls and logging mechanisms to track who accesses the system and how it is used. This includes maintaining logs of user actions and system interactions.

Third, CEASEFIRE has a full work package dedicated to the development of training. Users of the system will be required to undergo comprehensive training on ethical use, data privacy, and legal compliance.

Fourth, partners are iteratively monitoring the social impact of the project. In this regard, TRI is conducting a societal impact assessment. The mid-term version is reported in Section 5.

Finally, the project is monitored also by an independent ethics advisor who gives further advice to ensure CEASEFIRE adheres to ethical standards, legal requirements, and project goals.

3. Legal requirements

This section outlines the legal requirements for CEASEFIRE platforms applicable at the use phase.

3.1. Law Enforcement Directive

The table below outlines the main requirements that LEAs and tech partners should look at and make sure to comply with when CEASEFIRE is in the exploitation and use phase. National laws transposing the LED should also be considered since they might provide further guidance and further requirements.

It should be noted that CEASEFIRE consortium partners are taking measures to comply with the GDPR requirements during the research phase as at this stage, the LED does not apply to activities in the CEASEFIRE project.

Requirement	Obligation
Time-limits for storage and review Article 5	Specific time limits should be set for the erasure of personal data or for a periodic review of the need for storage of personal data. Procedural measures shall ensure that those time limits are observed. Specific time limits might be set out in national laws transposing the LED, sectoral legislation or by the relevant data protection authority.
Distinction between different categories of data subject Article 6	LEAs, where applicable and as far as possible, should make a clear distinction between personal data of different categories of data subjects: (a) Suspects (b) Persons convicted of a criminal offence (c) Victims (d) Other parties to a criminal offence, such as witnesses, persons who can provide information, contacts or associates of one of the persons mentioned in points (a) and (b)
Distinction between personal data and verification of quality of personal data Article 7	As far as possible, LEAs should distinguish between personal data based on facts from personal data based on personal assessments. LEAs shall take all reasonable steps to ensure that personal data which are inaccurate, incomplete or no longer up to date are not transmitted or made available.
Specific processing conditions Article 9	Personal data collected and processed shall not be processed for a purpose other than the investigation, detection, prevention, and prosecution of criminal offences unless authorised by law. In such cases, the GDPR will be applicable.
Data protection by design and by default Article 20	LEAs and technical partners shall implement appropriate technical and organisational measures to ensure privacy by design and by default, such as pseudonymization and data minimization. In particular, such measures shall ensure that by default personal data are not made accessible to an indefinite number of natural persons.
Processor Article 22	If LEAs were to purchase or sign licensing agreements with the tech partner to use the tools, different data protection roles should be assigned. If the technical partner in in the position of being a processor, the requirements of Article 22 (Article 28 of the GDPR) shall be met.

Logging Article 25	Logs shall be kept for when different users from LEAs collect, alter, consult, disclose, transfer, combine and erase personal data. The logs shall be used solely for verification of the lawfulness of processing, self-monitoring, ensuring the integrity and security of the personal data, and for criminal proceedings. LEAs and tech partners shall make the logs available to the supervisory authority on request
Data Protection Impact Assessment Article 27	Before using the tools, LEAs shall contact their DPOs and/or legal and data protection teams to assess whether a Data Protection Impact Assessment (‘**DPIA**’) should be carried out – when processing is likely to result in a high risk to the rights and freedoms of natural persons. DPIAs (if needed) shall be completed before LEAs use the tools.
Prior consultation of the supervisory authority Article 28	When (1) a DPIA has been carried out and it is indicated that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk; or (2) where using new tech involves a high risk. LEAs shall priorly consult with relevant data protection authorities before using the tool.
Security of processing Article 29	Controllers, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of the processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, in particular as regards the processing of special categories of personal data. Including but not limited to: (a) deny unauthorised persons access to processing equipment used for processing (‘equipment access control’); (b) prevent the unauthorised reading, copying, modification or removal of data media (‘data media control’); (c) prevent the unauthorised input of personal data and the unauthorised inspection, modification or deletion of stored personal data (‘storage control’); (d) prevent the use of automated processing systems by unauthorised persons using data communication equipment (‘user control’); (e) ensure that persons authorised to use an automated processing system have access only to the personal data covered by their access authorisation (‘data access control’); (f) ensure that it is possible to verify and establish the bodies to which personal data have been or may be transmitted or made available using data communication equipment (‘communication control’); (g) ensure that it is subsequently possible to verify and establish which personal data have been input into automated processing systems and when and by whom the personal data were input (‘input control’);

	(h) prevent the unauthorised reading, copying, modification or deletion of personal data during transfers of personal data or during transportation of data media ('transport control'); (i) ensure that installed systems may, in the case of interruption, be restored ('recovery'); (j) ensure that the functions of the system perform, that the appearance of faults in the functions is reported ('reliability') and that stored personal data cannot be corrupted by means of a malfunctioning of the system ('integrity').
Notification of data breaches Article 30	Controllers shall notify without undue delay and, where feasible, not later than 72 hours after having become aware of it, the personal data breach to the supervisory authority, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Processors shall notify the controller without undue delay after becoming aware of a personal data breach.
Data transfers Article 39	When personal data is transferred by competent authorities to a third country or to an international organisation, the transfer shall undergo the following conditions: • The transfer must be necessary for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security. • The data must be sent to an authority that deals with the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security. • If the data originally came from another country, that country must approve the transfer. • The European Commission must have decided that the receiving country provides adequate data protection, or there must be other safeguards in place, or special exceptions must apply. • If the data is being sent on to another country, the original transferring authority or another authority in the same country must approve it, considering factors like the seriousness of the crime involved and the data protection level in the receiving country. In urgent situations where there's an immediate and serious threat to public safety, data can be transferred without prior approval from another country if waiting for approval would take too long. The authority that would normally give approval must be informed right away.

Table 3 LED requirements.

3.2. AI Act

The AI Act [4] follows a risk-based approach, classifying AI systems as (1) prohibited, (2) high-risk and, (3) low-risk AI-systems. The AI Act will enter into force twenty days after its publication in the official Journal, and be fully applicable 24 months after its entry into force, with some exceptions. Bans on prohibited practises, will apply six months after the entry into force date; codes of practise which will apply nine months after entry into force, general-purpose AI rules including governance which will apply 12 months after entry into force, and obligations for high-risk systems after 36 months.

CEASEFIRE consists of a series of AI powered technologies and will likely fall within the scope of the forthcoming AI Act, once placed on the EU market and/or used in the EU [Art. 2(1)]. In its current form, the AI Act does not apply to any research, testing or development activity regarding AI systems or models prior to their being placed on the market or put into service [Art. 2(8)]. The testing of AI systems in real world conditions is not covered by the scientific research exemption of the AI Act. Nonetheless, it is most likely that the AI Act's obligations will still have a strong impact on AI research, considering the need to anticipate placement on the market or to test in real-world conditions.

The CEASEFIRE use cases and related tools would not fall under the category of prohibited systems Art. 5(1)] since CEASEFIRE tools are not:

- Deploying subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person's or a group of persons' behaviour by (…) impairing the person's ability to make an informed decision, causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person (…) significant harm.
- Exploiting any of the vulnerabilities of a person or a specific group of persons due to their age, disability or socio-economic situation with the objective to or the effect of materially distorting the behaviour of a person (…) pertaining to that group in a manner that causes (…) significant harm.
- Evaluating or classifying of natural persons or groups thereof over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics.
- Making risk assessments of natural persons in order to assess or predict the risk of a natural person to commit a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics.
- Creating or expanding facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage.
- Inferring emotions of a natural person in the areas of workplace and education institutions.
- Categorising individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement.
- Using 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement

AI systems referred to in Article 6 and in Annex III of the text shall be considered high-risk and will need to comply with the requirements set out in Title III. Point 6 of Annex III refers to law enforcement, and therefore the following systems would be categorised as high-risk within the law enforcement use:

- AI systems intended to be used (…) in support of law enforcement authorities or on their behalf to assess the risk of a natural person to become a victim of criminal offences;
- AI systems intended to be used (…) in support of law enforcement authorities as polygraphs and similar tools;
- AI systems intended to be used (…) in support of law enforcement authorities to evaluate the reliability of evidence in the course of investigation or prosecution of criminal offences;
- AI systems intended to be used (…) in support of law enforcement authorities for assessing the risk of a natural person of offending or re-offending not solely based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups;

- AI systems intended to be used (…) in support of law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences.

According to the Law Enforcement Directive Article 3(4) profiling is

> any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.

Considering the AI Act provisions, CEASEFIRE systems might fall under the category of high-risk AI systems. It's important to highlight that there are presently no established guidelines for interpreting the AI Act and that the research is ongoing. Therefore, the table and evaluation provided below **should not be viewed as conclusive.** The content might be subject to change. This implies that certain use cases and related tools could potentially be categorized differently in the future.

Use case	High risk [Yes/No/Difficult to assess]	Provision in Annex III and comments
Real-time systematic firearms incident and intelligence information collection and exchange	No	This use case might involve machine learning techniques to produce some outputs. It will not exhibit adaptiveness after deployment, but it will produce outputs as a response to some inputs. These outputs include: a) automatic extraction of standardised set of information from online news articles identified through queries, and b) analysis, red flags, risk indicators, and some predictions based on the data collected (e.g., incident type, date, location -country, city, coordinates-, firearm involved -category, type, model, brand-, and number of victims and perpetrators). The focus of the component will be on firearm incidents. All the analysis that will be produced will refer to events (i.e. firearm incidents). Natural persons will not be included in the analysis and will not be profiled. Based on the information above, the components of this use case might not be considered high risk.
On the spot firearm seizure registration and cross-border data search	Yes	This use case involves a series of AI algorithms used to build firearms trafficking networks. In principle, this tool might fall under point (f): AI systems used in support of law enforcement authorities for profiling of natural persons.
Firearms purchase on dark web marketplaces	Yes	This use case involves a variety of AI-based tools to identify patterns and correlations related to illicit firearms trafficking transactions made through Bitcoin or Ethereum. In principle, this tool might fall under point (f): AI systems used in support of law enforcement authorities for profiling of natural persons.

Mail order and courier service firearms trafficking detection using scanning technologies	Difficult to assess	This use case employs AI for firearms recognition. The technology does not involve any reliability scoring. However, this tool might fall under point "d) AI systems intended to be used (…) in support of law enforcement authorities to evaluate the reliability of evidence in the course of investigation or prosecution of criminal offences," depending on the interpretation that jurisprudence assigns to the notion of evaluating evidence reliability. Guidance should come from the EC AI board in this regard.
3D printed firearm blueprints distribution	Yes	The AI tool that is part of this use case is aimed at identifying and categorising discussions related to gun trafficking and 3D printing of firearms. Partners are developing a 1) Suspiciousness Model to differentiate potentially suspicious conversations from the larger pool of discussions related to firearms and an 2) Intent Recognition model, which classifies conversation individual messages based on the user's purpose into eight categories: advice, offer, request, exchange, tutorial, social, information-seeking and comment. In principle, this tool might fall under point (f): AI systems used in support of law enforcement authorities for profiling of natural persons.

Table 4 High-risk AI tools assessment

Even if CEASFIRE does not end up falling within the high-risk category, obligations that apply to high-risk AI systems, which are summarised in the tables below, serve as concrete steps to comply with the general principles for trustworthy AI.

3.2.1.1. Requirements for high-risk systems

The AI Act outlines specific obligations for providers (technical partners), distributors and end-users of high-risk systems. The tables below provide a summary of the obligations around high-risk AI systems. Technical partners should aim at complying with these requirements so their tools can ensure safety and respect for fundamental rights throughout the entire lifecycle of an AI system. In addition, by following these requirements the tools will not become obsolete when the AI Act, comes into force.

Article 3 (3) of the AI Act defines "providers" as "*a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge*" According to this definition, the technical partners of CEASEFIRE shall be considered providers. The below table presents an overview of the obligations for technical partners during the development and use phase of high-risk AI tools.

Obligations of providers
Chapter III, Section 3

Obligations of providers Article 16	• Ensure compliance with Section 2 (Chapter III) requirements. • Indicate name, registered trade name or trade mark, address on the packaging or accompanying documentation. • Have a quality management system as per Article 17. • Keep the documentation as per Article 18. • Keep the logs automatically generated as per Article 19. • Ensure the system undergoes conformity assessment as per Article 43. • Draw up EU declaration of Conformity as per Article 47. • Affix the CE marking as per Article 48. • Comply with registration obligation as per article 49(1). • Take corrective actions as per Article 20. • Upon reasoned request of a competent authority, demonstrate conformity with the requirements of Section 2 (Chapter III). Ensure compliance with accessibility requirements, in accordance with Directive 2019/882 on accessibility requirements for products and services and Directive 2016/2102 on the accessibility of the websites and mobile applications of public sector bodies.
Quality Management system Article 17	Providers of high-risk AI systems shall put a quality management system in place that ensures compliance with this Regulation. That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions, Aspects that should be included are detailed in Article 17. Note that, as per Article 63, microenterprises may fulfil certain elements of the quality management system in a simplified manner (a microenterprise is defined as an enterprise which employs fewer than 10 persons and whose annual turnover and/or annual balance sheet total does not exceed EUR 2 million.)
Documentation keeping Article 18	Providers shall keep for a period ending 10 years after the AI system has been placed on the market or put into service: - Technical documentation (Article 11) ; - Quality management system documentation (Article 17) ; - Changes approved by notified bodies, where applicable; - Decision and other documents issued by notified bodies, where applicable; - EU declaration of conformity (referred to in Article 47).

Automatically generated logs Article 19	Obligation to keep the logs automatically generated by high-risk AI systems, to the extent that such logs are under the provider's control by virtue of a contractual arrangement with the user or otherwise by law. Logs must be kept for a period of at least 6 months, unless provided otherwise in applicable EU or national law, especially in EU law on the protection of personal data.
Corrective actions and duty of information Article 20	If a high-risk AI system placed on the marked or put into service is not in conformity with this AI Act, providers shall put in place corrective measures: bring into conformity, withdraw, or recall the AI system, as appropriate. They shall inform the distributors of the high-risk AI system in question and, where applicable, the deployers, the authorised representative and importers accordingly.
Cooperation with competent authorities Article 21	Duty to provide all information and documentation necessary to demonstrate the conformity of the high-risk AI system with the requirements set out in Section 2 (Chapter III) to competent authorities.
Authorised representatives Article 22	Prior to making their systems available on the Union market providers established outside the Union shall, by written mandate, appoint an authorised representative which is established in the Union.
Chapter III, Section 5	
Conformity assessment Article 43	Providers shall follow the conformity assessment procedure based on internal control as referred to in Annex VI, which does not provide for the involvement of a notified body.
EU declaration of conformity Article 47	Draw up a written machine readable, physical or electronically signed EU declaration of conformity for each high-risk AI system and keep it at the disposal of the national competent authorities for 10 years after the AI high risk system has been placed on the market or put into service. The EU declaration of conformity shall state that the high risk AI system meets the requirements set out in Section 2 (Chapter III).
CE marking of conformity Article 48	For high-risk AI systems provided digitally, a digital CE marking shall be used, only if it can be easily accessed via the interface from which the AI system is accessed or via an easily accessible machine-readable code or other electronic means.
Registration of high-risk AI systems in EU database Article 49	Before placing on the market or putting into service a high-risk AI system, the provider or, where applicable, the authorised representative shall register themselves and their system in the EU database referred to in Article 71.
Chapter VI	

Testing of high-risk AI systems in real world conditions outside AI regulatory sandboxes Article 60	Testing of AI systems in real world conditions outside AI regulatory sandboxes may be conducted by providers or prospective providers of high-risk, in accordance with the provisions of this Article and the real-world testing plan referred to in this Article, without prejudice to the prohibitions under Article 5.
Informed consent Article 61	For the purpose of testing in real world conditions, freely-given informed consent shall be obtained from the subjects of testing prior to their participation in such testing and after their having been duly informed with concise, clear, relevant, and understandable information regarding: the nature and objectives of the testing, conditions, rights of participants.
Chapter IX, Section 1 and 2	
Post market monitoring Article 72	Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the AI technologies and the risks of the high-risk AI system.
Reporting serious incidents Article 73	Providers of high-risk AI systems placed on the Union market shall report any serious incident to the market surveillance authorities of the Member States where that incident occurred.

Table 5 Obligations of providers.

The table below provides a specification of the requirements for high-risk systems.

Requirements for high-risk systems	
Chapter III, Section 2	
Risk management system Article 9	A risk management system must be established, implemented, documented and maintained in relation to high-risk AI systems. It shall comprise a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating. It will include: (a) identification and analysis of the known and the reasonably foreseeable risks that the high-risk AI system can pose to the health, safety or fundamental rights when the high-risk AI system is used in accordance with its intended purpose; (b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse; (c) Adoption of appropriate and targeted risk management measures designed to address the risks identified. The risks referred to above only concern those which may be reasonably mitigated or eliminated through the development or

	design of the high-risk AI system or the provision of adequate technical information. The ultimate goal is that the residual risk posed by the high-risk AI system is judged acceptable. Guidance for the risk management system can be found in Article 9(a)-(c).
Data and Data governance Article 10	Training, validation and testing data sets shall be subject to appropriate data governance and management practices. These practices include: design choices, data collection and processing procedures, the formulation of assumptions, an assessment of the suitability of dataset, an examination of possible bias and appropriate measures to detect and mitigate it, the identification of gaps that prevent compliance with the regulation. Data sets shall: • Be relevant, representative and, to the best extent possible, free of errors and complete. • Have the statistical properties of the persons or groups of persons in relation to whom the high risk AI system is intended to be used • Take into account, to the extent required by the intended purpose, characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used. • Include the processing of special categories of personal data (according to Article 9(1) GDPR, Article 10 LED or Article 10(1) of EUDPR) only to the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to high-risk AI systems.
Technical Documentation Article 11	Technical documentation pursuant to Annex IV must be in place before the high-risk AI system is placed on the market or put into service. It must be kept up-to date and shall demonstrate compliance with the high-risk requirements set out in the Regulation.
Record Keeping Article 12	The system should allow for the automatic recording of events (logs). The logs should ensure a level of traceability of the AI system's functioning. The logging capabilities should provide: • Recording of the period of each use of the system • Reference database against which input data has been checked by the system • Input data for which the search has led to a match • Identification of natural persons involved in the verification of the results as per Article 14 (5). Important: Logging capabilities shall enable the monitoring of the operation of the high-risk AI system with respect to the occurrence

	of situations that may result in the AI system presenting a risk within the meaning of Article 65(1).
Transparency and provision of information to deployers Article 13	The AI system shall be developed in such a way to ensure their operation is sufficiently transparent to enable deployers to interpret the system's output and use it appropriately. To that extent, the systems shall be accompanied by instructions for use in appropriate digital format. Specifying: a) the identity and contact details of the providers, b) characteristics, capabilities and limitation of performance of the system, c) changes to the system and its performance d) the human oversight measures referred to in Article 14, e) the computational and hardware resources needed, the expected lifetime of the high risk AI system and any necessary maintenance and care measures, ea) a description of the mechanisms included within the AI system that allows users to properly collect, store and interpret the logs.
Human oversight Article 14	High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use. The goal is to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Oversight measures can be ensured through a) system design and b) other measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.
Accuracy, robustness and cybersecurity Article 15	The levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use. The robustness of high-risk AI systems may be achieved through technical redundancy solutions, which may include backup or fail-safe plans. High-risk AI systems that continue to learn after being placed on the market or put into service must be developed in a way to ensure include that possibly biased outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures.

Table 6 Requirements for high-risk systems.

The AI Act outlines obligations of distributors and employers. According to Article 3 (7) of the AI Act, "distributor" means "*any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market*".

Obligations of distributors	
Obligations Article 24	Verify that the high-risk AI system bears the required CE conformity marking, that it is accompanied by a copy of EU declaration of conformity and instruction of use, and that the provider and the importer of the system, as applicable, have complied with their obligations set out in Article 16, point (aa) and (b) and 26(3) respectively. Shall not make the high-risk AI system available on the market until that system has been brought into conformity with those requirements. If there is a reason to consider that the system is not in conformity with the requirements of Section 2 (Chapter III), take the corrective actions necessary to bring the system to conformity with those requirements. Provide authority with all the information and documentation upon a reasoned request. Cooperate with national competent authorities on action those authorities take to reduce or mitigate the risk posed by the high risk AI system.

Table 7 Obligations of distributors.

Article 3 (4) of the AI Act defines ‘deployers’ as *“any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity”*. In CEASEFIRE the end-users are LEAs. The table below summarises some of the obligations that end-users of high-risk AI systems shall comply with when acquiring or using CEASEFIRE tools.

Obligations of deployers of high-risk AI systems	
Obligations Article 26	Deployers shall take appropriate technical and organisational measures to ensure they use such systems in accordance with the instructions of use accompanying the systems
	Deployers shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support.
	Deployers shall ensure that the natural persons assigned to ensure human oversight of the high-risk AI systems have the necessary competence, training and authority as well as the necessary support.
	To the extent users exercises control over the input data, users shall ensure that input data is relevant and sufficiently representative in view of the intended purpose of the high-risk AI system.
	Deployers shall monitor the operation of the high-risk AI system on the basis of the instructions of use and when relevant, inform

	providers in accordance with Article 72. When they have reasons to consider that the use in accordance with the instructions of use may result in the AI system presenting a risk within the meaning of Article 79(1) they shall, without undue delay, inform the provider or distributor and relevant market surveillance authority and suspend the use of the system. They shall also immediately inform first the provider, and then the importer or distributor and relevant market surveillance authorities when they have identified any serious incident If the deployer is not able to reach the provider, Article 73 shall apply mutatis mutandis. This obligation shall not cover sensitive operational data of users of AI systems which are law enforcement authorities.
	Deployers shall keep the logs automatically generated by that high-risk AI system to the extent such logs are under their control for a period appropriate to the intended purpose of the high-risk AI system, of at least six months, unless provided otherwise in applicable Union or national law, in particular in Union law on the protection of personal data.
	Deployers of high-risk AI systems that are public authorities or Union institutions, bodies, offices and agencies shall comply with the registration obligations referred to in Article 49. When they find that the system that they envisage to use has not been registered in the EU database referred to in Article 71 they shall not use that system and shall inform the provider or the distributor.
Fundamental rights impact assessment **Article 27**	Prior to the first use of a high-risk AI system users shall perform an assessment of the impact on fundamental rights that the use of the system may produce. For that purpose, users shall perform an assessment consisting of: a) a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose; b) a description of the period of time and frequency in which each high-risk AI system is intended to be used; c) the categories of natural persons and groups likely to be affected by its use in the specific context; d) the specific risks of harm likely to impact the categories of persons or group of persons identified pursuant point (c), taking into account the information given by the provider pursuant to article 13; e) a description of the implementation of human oversight measures, according to the instructions of use; f) the measures to be taken in case of the materialization of these risks, including their arrangements for internal governance and complaint mechanisms.

	If the system has been already used by others, the user may, in similar cases, rely on previously conducted fundamental rights impact assessments or existing impact assessments carried out by provider. If, during the use of the high-risk AI system, the deployer considers that any of the factors are or no longer up to date, the user will take the necessary steps to update the information. Once the impact assessment has been performed, the user shall notify the market surveillance authority of the results of the assessment. If these obligations are already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment be conducted in conjunction with that data protection impact assessment.

Table 8 Obligations of deployers.

4. Requirements for training materials

This section outlines some key requirements that TRI has identified for CEASEFIRE training material:

- Clear and separate information for each specific functionality regarding:
    a) intended purpose
    b) the logic behind the system
    c) limitations and benefits
    d) description on the information used (data, inputs)
    e) accuracy levels and any other relevant metrics
    f) the key design choices including the rationale and assumptions
    g) description of the system architecture
    h) training methodologies
    i) validation and testing procedures used
    j) information about the monitoring, functioning and control of the AI system.

- Clear instructions on how to interact with the interfaces.
- Basic information on statistic and probability concepts that are useful to interpret the system output.
- Guidance on how the results should be interpreted and critical thinking.
- Basic guidance on data protection.
- Basic guidance on trustworthy AI.
- Ethics training on responsible use of the crawler.
- Information on misuse.
- Information on automation bias.

5. Interim assessment of societal impacts

This section presents the interim societal impact assessment of CEASEFIRE. Based on the ASSERT[1] methodology presented in D1.4, we performed an analysis of the three dimensions, namely: i) how the project meets the needs of society; ii) how the project could have potential negative impacts; and iii) how the project could potentially benefit society. For each dimension, we answered the set of questions provided by the ASSERT methodology. The results are reflected in the following tables.

The ASSERT methodology provides a set of questions to guide the evaluation of the societal impact of security research and innovation. The ASSERT methodology is particularly suited to assess CEASEFIRE impacts for several reasons. First, the ASSERT approach is designed for security research and innovative application of security technologies. Second, the approach has a participatory nature and requires engagement of different stakeholders at an early stage of the project. Third, it intends the societal impact assessment activity as an ongoing process of analysing and monitoring the intended and unintended social consequences of innovations.

Assessment of the first dimension: Ensuring the research project meets the needs of society	
Question	**Assessment**
Which documented societal security need(s) does the proposed research address?	Firearms have long been identified by the EU as a major threat for citizens. They can increase the danger posed by serious and organised crime, including drug trafficking, human trafficking, violent crime and terrorism [1]. In 2017, it was estimated that civilians in the EU possessed around 35 million illicit firearms, accounting for 56% of the total estimated firearms [5]. The 2020-2025 EU action plan on firearms trafficking calls for: • building a better intelligence picture; • enabling simultaneous searches in different firearms datasets; • a better monitoring of illicit trades the dark web Marketplace.
How will the research output meet these needs? How will this be demonstrated? How will the level of societal acceptance be assessed?	CEASEFIRE addresses these needs by providing a set of technologies that will allow to: • Perform simultaneous searches in several national and international databases. • Automatically recognise firearms. • Merge online and offline data regarding firearms, their critical components and blueprints to build firearms trafficking networks and integrated reports. CEASEFIRE functionalities will be tested through dedicated pilots. Societal acceptance is one of the elements considered in the present impact assessment.

[1] The ASSERT project, which was co-funded by the European Commission under the 7th Framework Programme call FP7-SEC-2012-2, work programme topic 6.3.2, “Criteria for assessing and mainstreaming societal impacts of EU security research activities – Coordination and support action”.

Is the research project aware of challenges to these needs?	Up to M18, we identified several challenges related to these needs: • The accessibility of online and offline information. • The effective exchange of information on firearms trafficking between countries. • The technical capabilities of LEAs' staff.
Does addressing the documented societal needs through the proposed research require any trade-offs with other documented societal needs? How is this trade-off decided? Is this trade-off still valid if the research is less effective than anticipated?	Some of the data that CEASEFIRE will analyse are acquired through a crawler that automatically collects surface as well as dark web data. By their nature, web crawlers might pose significant concerns regarding the privacy of surface and dark web users. During the research phase, CEASEFIRE partners are mitigating these risks by 1) carrying out a highly targeted crawling activity that is limited only to firearms related keywords, 2) anonymising and pseudonymising personal data, and 3) not handling raw data that could contain personal information. In order to mitigate these risks at a use phase, training will be provided to users on the crawler capabilities and limitations, risks of misuse and ethical use of the crawler. Accuracy levels of the algorithms used are monitored during the development process. Regular audits to monitor the system performance will be suggested to users. It is important to note that CEASEFIRE technologies are designed to respond to firearms trafficking threats and incidents after they have occurred, rather than to prevent or detect them beforehand. This means that CEASEFIRE technologies are used to analyse and investigate incidents of firearms trafficking post-occurrence, helping to understand how the trafficking took place. If the research is less effective than anticipated, the findings will provide a solid background for improvement through future research.
What threats to society does the research address? (e.g., crime, terrorism, pandemic, natural and man-made disasters).	CEASEFIRE aims to help LEAs to better monitor, analyse, understand firearms trafficking activities. Given the correlation between firearms trafficking and other types of organised crime, CEASEFIRE outcomes will also indirectly contribute to address: • Violent crime • Human trafficking • Drug trafficking • Terrorism.
How is the proposed research appropriate to address these threats?	The research is appropriate to address these threats because the outcomes would allow LEAs to: • have a comprehensive picture of online and offline firearms trafficking activities related to specific incidents and investigation leads and understand how different firearms related events might be associated with one another. • Perform simultaneous searches across different databases. • Automatically recognise firearms in parcels; recognise the make, model and calibre of a firearm through a mobile application. All tools are being developed following a compliance by design and human-centric approach.

	Ethical and legal risks are being monitored throughout the project lifecycle and mitigation measures are being proposed and implemented by all partners. End-users' perspectives and needs are being integrated in the technology design.
7. What other measures could be adopted to address these threats?	Firearms trafficking is a complex international phenomenon. Standardisation of firearms names, as well as reporting protocols could be of help for database searchers. Technical training for LEAs in emerging technologies is also key to help officers to understand possibilities and actively shape the future of their work.

Table 9 Ensuring the project meets the needs of society.

Assessment of second dimension: **Ensuring the research project does not have negative impacts on society**		
Question	**Assessment**	**Mitigation**
How could the project have a negative impact on fundamental rights and freedoms?	The data collected could potentially contribute to reinforce LEA bias toward certain social groups, if found to be more often implicated in firearms trafficking activities. Furthermore, the crawling activity might have a negative impact on the right to privacy.	Strategies for bias prevention and mitigation are implemented in algorithm design and data handling. State-of-the-art technical tools are utilized for a comprehensive understanding of data, model, and performance (see the bias and discrimination section of each task assessment for more information). The dedicated ethics tasks in WP1 and WP10 undertake systematic assessment of technical project activities and advise on tactics to mitigate risks and avoid negative association stemming from specific parts of the datasets. Users will be given ethical training on data interpretation and will encourage officers to integrate the CEASEFIRE output with additional contextual information. Users will be advised to regularly audit the system to monitor its performance. Ethics and compliance by design is embedded in the technology development phase. Partners are undertaking several measures to embed privacy in the design of the technology such as 1) not processing raw data, 2) pseudonymising and anonymising data (see the privacy and data governance section).
How could the project have negative impacts on the right to life?	The project is not expected to have a negative impact on the right to life.	N/A
How could the research have a negative impact on privacy?	As the project is collecting crawler-based online information, there is a potential for negative impact on privacy, through the collection of personal data posted online.	The crawler data collection, as well as its labelling and use by various project activities are systematically monitored by the ethical assessment. Ad hoc mitigation measures are being implemented in each task (see the privacy and data governance section of each task assessment for more information). Sensitivisation activities with end-users on the use

		of personal data in line with the principles of minimisation and proportionality are considered for the training activities.
How could the research have a negative impact on freedom of movement and assembly and on access to public spaces?	It is not expected that freedom of movement /assembly and access to public spaces would be affected within the remit of the project.	N/A
How could the research have a negative impact on working conditions?	Because of organisational pressure, users might be pushed to rely solely on the output of the system without further contextualisation or investigation. The use of CEASFIRE would require the adoption of new skills by users regarding basic statistical and probability concepts. It would also require ethical training on the output interpretation.	Users are and will be trained on output interpretation and the importance of using extra information to contextualise and further investigate the information given by the system. The training material will contain information on basic statistical and probability concepts.
How could the research have a negative impact on the principle of democracy?	Although the project aims to facilitate social justice, unfair assumptions or treatment of specific social groups might potentially negatively affect democracy.	During the course of the project, sensitisation workshops have targeted LEAs in an effort to demonstrate the risks of overreliance on large-scale datasets, including the formation of biases, prejudices and unfair assumptions. The final training material will contain 1) information on how to mitigate risks of bias and discrimination when using the system and interpreting data and 2) information on how to use the crawler ethically.
If implemented, how could the research have a negative impact on this aspect (culture and community, way of life, etc.)?	There could be a significant potential of impacting certain social groups (e.g., ethnicities and nationalities asymmetrically represented in the CEASEFIRE datasets).	Strategies for bias prevention and mitigation are implemented in algorithm design and data handling. State-of-the-art technical tools are utilized for a comprehensive understanding of data, model, and performance (see the diversity, non-discrimination and fairness section).
How could the research impact disproportionately upon specific groups or unduly discriminate against them? How could the research increase discrimination?	As described above, the data collected could potentially inform LEA bias of certain social groups, e.g., nationals of countries more often represented in the datasets as involved in firearms trafficking activities.	Strategies for bias prevention and mitigation are implemented in algorithm design and data handling. State-of-the-art technical tools are utilized for a comprehensive understanding of data, model, and performance (see

		the diversity, non-discrimination and fairness section).
Could the research have impacts upon vulnerable groups?	CEASEFIRE may negatively impact social groups that are already stigmatised as being intrinsically associated with organised crime or terrorism (for instance, people who have been convicted, ex-prisoners, persons pertaining to sub-cultures, those from certain ethnic or religious minorities.). The possibilities of such risks occurring would depend on a range of factors, namely: i) the quality of the data for training the algorithms; ii) how the models are validated and tested; and iii) how end-users make decisions based on the output of the system.	Strategies for bias prevention and mitigation are implemented in algorithm design and data handling. State-of-the-art technical tools are utilized for a comprehensive understanding of data, model, and performance (see the diversity, non-discrimination and fairness section). In this regard, another risk mitigation technique is the provision of training and understanding on bias and discrimination (amongst other concerns) that may materialise in the use of the technology, and how to mitigate them.

Table 10 Ensuring the research project does not have negative impacts on society.

Assessment of third dimension: **Ensuring the research project benefits the society**	
Question	**Assessment**
What segment(s) of society will benefit from increased security as a result of the proposed research? How will they benefit?	The most glaring benefit of the project will be to law enforcement and judicial authorities in Europe, as its outputs aim to facilitate a better information picture and evidence collection in the investigation of firearms trafficking and other firearms-related criminal activities.
Are additional measures required to achieve this benefit?	LEA organisations need to implement the CEASEFIRE systems in their practices, considering data management policies, and clear and transparent guidelines.
Are additional measures possible to extend these benefits to other segments of society?	More transparency and dissemination measures to guarantee that citizens are aware of the use of this technology.
In what contexts might this benefit be lacking or not be delivered by the research project?	Not applicable.
How will society as a whole benefit from the proposed research?	With a better information picture and evidence collection in the investigation of firearms trafficking and other firearms-related criminal activities, LEAs should be better able to predict and prevent firearms-related incidents, which would affect society at large, by providing enhanced security to citizens. CEASEFIRE has the potential to reduce the illicit exchange of firearms and disrupt the financial flow within organised crime groups that sustain their activities through firearms trafficking e.g., [1]. By aiding in the identification of criminal networks and the prevention of illegal firearms trade, CEASEFIRE will help uphold the rule of law, which (in addition to public safety), is a fundamental pillar of democracy.
Are there other European societal values that are enhanced by the research (e.g., public accountability and transparency; strengthened community engagement, human dignity; good governance; social and territorial cohesion; sustainable development.)	The output of CEASEFIRE has the potential of affecting various European societal values beyond the obvious benefit of enhanced security. Accountability and transparency of the investigation and prosecution of criminality related to firearms would be increased by the improved information picture, which in turn could contribute to more efficient decision making. The logging capabilities of CEASEFIRE would also contribute toward increased transparency of the investigation process. The fact that the applications are designed to make cross-border investigations more effective will aid territorial cohesion among various regions within Europe, including those with less ample financial and human resources at LEAs' disposal. In addition, international cooperation within the EU will contribute to a safer EU community, while cooperation with non-EU partners contributes to the EU's neighbourhood policy as it enhances the EU's partnerships on topics of particular interest (such as firearms trafficking). Finally, as the applications are intended for use by any LEA, this contributes to sustainability of effort, as it pre-empts the need for each national authority to develop their own systems.

Table 11 Ensuring the research project benefits society.

6. Conclusion

This deliverable has provided an overview of how CEASEFIRE is aligning its technologies with the 7 principles for ethical and trustworthy AI. At this stage of the project (M18) the technologies have not been finalised yet, however after the ethical assessment it can be concluded that they are in a good position with regards to achieving ethical and trustworthy AI. Furthermore, this deliverable presented an overview of the legal requirements under the LED and the AI Act, which the tools shall comply with if the product will achieve the use phase. Finally, this deliverable presented an updated version of the societal impact assessment. As research tasks are ongoing and tools are not yet finalised, the content presented in this deliverable can be subject to variation. A final version (D1.6) with the updates will be delivered in in M36.

7. References

[1] UNODC, “Global Study on Firearms Trafficking 2020,” United Nations Office on Drugs and Crime. [Online]. Available: https://www.unodc.org/documents/data-and-analysis/Firearms/2020_REPORT_Global_Study_on_Firearms_Trafficking_2020_web.pdf

[2] Europol, “EU Policy Cycle - EMPACT,” 2023. [Online]. Available: https://www.europol.europa.eu/crime-areas-and-trends/eu-policy-cycle-empact

[3] AI HLEG, “Ethics Guidelines for Trustworthy AI,” European Commission, 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[4] European Parliament, “European Parliament legislative resolution of 13 March 2024 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 (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).” 2024. [Online]. Available: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

[5] European Commission, “COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS 2020-2025 EU action plan on firearms trafficking,” 2020, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0608&from=nl.

[6] D. R. Becker, C. C. Harris, W. J. McLaughlin, and E. A. Nielsen, “A participatory approach to social impact assessment: the interactive community forum,” *Environ. Impact Assess. Rev.*, vol. 23, no. 3, pp. 367–382, May 2003, doi: 10.1016/S0195-9255(02)00098-7.